THE UNIVERSITY
of EDINBURGH

The generalised Langevin equation: asymptotic properties and numerical analysis

Matthias Sachs

Doctor of Philosophy
University of Edinburgh
2017

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(*Matthias Sachs*)

To my grandparents

Lay summary

The generalised Langevin equation (GLE) is a non-Markovian stochastic differential equation widely used in model reduction of complicated nonlinear dynamical systems, where, for example, it describes the dynamics of a molecular system in contact with an active environment. When the GLE is discretised it provides a sequence of samples approximately drawn form the invariant distribution. Thus the discretised GLE can be viewed as either a method for modelling dynamical response of a subsystem embedded in a complex medium or as an alternative to Markov Chain Monte Carlo (MCMC) methods which are among the most widely used algorithms in all of computational and data science. While ergodic properties, numerical analysis and the construction of discretisation schemes for the underdamped and overdamped Langevin equation are well established, only limited work has been done thus far in the case of the generalised Langevin equation.

Abstract

In this thesis we concentrate on instances of the GLE which can be represented in a Markovian form in an extended phase space. We extend previous results on the geometric ergodicity of this class of GLEs using Lyapunov techniques, which allows us to conclude ergodicity for a large class of GLEs relevant to molecular dynamics applications. The main body of this thesis concerns the numerical discretisation of the GLE in the extended phase space representation. We generalise numerical discretisation schemes which have been previously proposed for the underdamped Langevin equation and which are based on a decomposition of the vector field into a Hamiltonian part and a linear SDE. Certain desirable properties regarding the accuracy of configurational averages of these schemes are inherited in the GLE context. We also rigorously prove geometric ergodicity on bounded domains by showing that a uniform minorisation condition and a uniform Lyapunov condition are satisfied for sufficiently small timestep size. We show that the discretisation schemes which we propose behave consistently in the white noise and overdamped limits, hence we provide a family of universal integrators for Langevin dynamics. Finally, we consider multiple-time stepping schemes making use of a decomposition of the fluctuation-dissipation term into a reversible and non-reversible part. These schemes are designed to efficiently integrate instances of the GLE whose Markovian representation involves a high number of auxiliary variables or a configuration dependent fluctuation-dissipation term. We also consider an application of dynamics based on the GLE in the context of large scale Bayesian inference as an extension of previously proposed adaptive thermostat methods. In these methods the gradient of the log posterior density is only evaluated on a subset (minibatch) of the whole dataset, which is randomly selected at each timestep. Incorporating a memory kernel in the adaptive thermostat formulation ensures that time-correlated gradient noise is dissipated in accordance with the fluctuation-dissipation theorem. This allows us to relax the requirement of using i.i.d. minibatches, and explore a variety of minibatch sampling approaches.

Acknowledgments

First and foremost I want to thank my advisor Benedict Leimkuhler. His perspective on mathematics and science and the role a numerical analysts plays in the interface between these two fields of research has shaped my thinking in my ongoing attempt of becoming a scientist and I wouldn't want to miss the insight into these fields which I gained through his supervision. Similarly, I want to thank my co-advisor Vincent Danos who as well has been extremely supportive throughout my PhD studies and who provided me yet with another perspective on mathematical modelling which I wouldn't want to miss. Working with both my advisors has made my PhD studies an extremely exciting and enjoyable experience.

I want to thank Gabriel Stoltz who in the course of my final year of my PhD studies has helped me to extend my knowledge in stochastic analysis and who patiently responded to all my questions I sent to him via email. I also want to thank my examiners Benjamin Goddard and Grigorios A. Pavliotis whose comments and critique helped to substantially improve the presentation and content of the revised version of this thesis. I thank Konstantinos C. Zygalakis for very helpful discussions and I am grateful to Mark Rowland, Jianfeng Lu, and Jonathan Mattingly for their contributions to this thesis which are detailed within the main body. Finally, I want to thank Ralf Banish, Nicolas Behr, Andrew Duncan, Gianpaolo Gobbo, Ricardo Honorato-Zimmer, Anton Martinsson, Charles Matthews, Zofia Trstanova, William Waites, and everyone else who at the moment of writing this acknowledgement section I might have accidentally missed to list here, for many stimulating and insightful discussions.

Throughout my PhD studies I was supported by ERC project RULE (grant number 320823).

Contents

Chapter 1

Introduction

1.1 Molecular dynamics

The fundamental modelling principle of molecular simulations is the abstraction of matter as a collection of point particles, which (at least in the traditional formulation of molecular dynamics) obey the laws of classical mechanics, i.e., the interaction between such particles is described by a force field and the evolution in time is governed by Newton's equations of motion. While this approach provides a microscopic perspective on matter, the interest typically lies in retrieving macroscopic properties of the simulated system, e.g., thermodynamic properties such as pressure as a function of density, or dynamical properties such as transport coefficients or rare event frequencies. A consequence of the statistical mechanical perspective is that both in the formulation of molecular dynamics models and in the numerical implementation of these, the focus lies on the accurate reproduction of macroscopic quantities rather than on obtaining trajectories, which at the a microscopic level satisfy the laws of classical mechanics. This approach is in strong contrast to the way simulation of particle systems (or more general N-body problems) is approached in other fields of classical mechanics such as in astronomy where in many cases the micro-description, e.g., the exact replication of trajectories of planetary systems, is the goal of study.

In the remainder of this section we will briefly review different flavours of molecular dynamics. We start by describing the setup of classical, i.e., Hamiltonian, molecular dynamics and show how the above-described macroscopic perspective of the dynamical systems naturally motivates the introduction of thermostats. We introduce the generalised Langevin equation at the end of this section as a more flexible version of the underdamped Langevin equation.

1.1.1 Notation and system description

Throughout this thesis the generic system we work with is a collection of point particles, i.e., the state of each particle is fully characterised by its position and momentum vectors. We denote by N the number of particles and by d the spatial dimension, so that $n = d \cdot N$ is the total number of configurational degrees of freedom. We further denote by $\boldsymbol{m}_i, i \in \{0, \dots, N\}$, the masses of the particles, by $\boldsymbol{p} = (\boldsymbol{p}_1, \dots, \boldsymbol{p}_{dN}) \in \mathbb{R}^n (=: \Omega_{\boldsymbol{p}})$ the momentum vector of all particles and by $\boldsymbol{q} = (\boldsymbol{q}_1, \dots, \boldsymbol{q}_n) \in \Omega_{\boldsymbol{q}}$, the vector of positions of all particles. Typically, particles are simulated either on an unbounded domain, in which case $\boldsymbol{q}_i \in \mathbb{R}$, hence $\Omega_{\boldsymbol{q}} = \mathbb{R}^n$, or in a periodic box, in which case $\boldsymbol{q}_i \in L\mathbb{T}$, hence $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n$, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denotes the one-dimensional unit torus and $L > 0$ the length of the simulation box.

The force field is typically given as the (negative) gradient of a potential energy function $U : \Omega_{\boldsymbol{q}} \to \mathbb{R}$, which is required to be at least almost everywhere continuously differentiable. The potential energy function is in practice either directly derived from first principles, i.e. from a quantum mechanical model (See e.g. [71] or [38]) or is constructed as an empirical approximation of the former. As the particles are viewed as point masses, the collective kinetic energy of the system is

$$E_{kin}(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T \boldsymbol{M}^{-1}\boldsymbol{p},$$

and hence the total energy/Hamiltonian of the particle system takes the simple form

$$H(\boldsymbol{q}, \boldsymbol{p}) = U(\boldsymbol{q}) + E_{kin}(\boldsymbol{p}). \tag{1.1}$$

1.1.2 Hamiltonian dynamics and symplectic integration

As the particles, representing the nuclei of atoms, in molecular dynamics are assumed to obey the laws of classical mechanics the time evolution of the particle system is described by Newton's equations of motion, i.e.,

$$\begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{M}^{-1}\boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}), \end{aligned} \tag{1.2}$$

where $\boldsymbol{q}, \boldsymbol{p}$ and U are as described in Section 1.1.1 and

$$\boldsymbol{M} = \operatorname{diag}(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_N) \otimes \boldsymbol{I}_d,$$

with $\boldsymbol{I}_d$ denoting the identity matrix in $\mathbb{R}^d$.

Let $\Phi_t : \Omega_{\boldsymbol{q}} \times \mathbb{R}^n \to \Omega_{\boldsymbol{q}} \times \mathbb{R}^n$, denote the flow map associated with (1.2), which is defined such that

$$\Phi_t(\boldsymbol{q}(0), \boldsymbol{p}(0)) = (\boldsymbol{q}(t), \boldsymbol{p}(t)),$$

for any initialisation $(\boldsymbol{q}(0), \boldsymbol{p}(0)) \in \Omega_{\boldsymbol{q}} \times \mathbb{R}^n$. We recall that the following properties hold for Φ_t in the case of Newton's equations (see e.g. [73]):

(i) the collection of flow maps $(\Phi_t)_{t\in\mathbb{R}}$ forms an algebraic group in the sense that

$$\Phi_t \circ \Phi_s = \Phi_{t+s},$$

with neutral element $\Phi_0 = \mathrm{Id}$, and well defined inverse

$$\Phi_t^{-1} = \Phi_{-t}.$$

(ii) symmetry holds in the sense that

$$S \circ \Phi_t \circ S = \Phi_{-t},$$

where the function $S : (\boldsymbol{q}, \boldsymbol{p}) \mapsto (\boldsymbol{q}, -\boldsymbol{p})$, reverses (flips) the momentum.

(iii) total energy is conserved under Φ_t:

$$H(\boldsymbol{q}(t), \boldsymbol{p}(t)) = H(\boldsymbol{q}(0), \boldsymbol{p}(0)),$$

(iv) in the absence of an external force, or more generally, if $\sum_{i=1}^{n} \partial_{\boldsymbol{q}_i} U(\boldsymbol{q}) = 0$, total momentum is conserved under Φ_t:

$$\sum_{i=1}^{N} \boldsymbol{p}_i(0) = \sum_{i=1}^{N} \boldsymbol{p}_i(t).$$

Besides the elementary conservation properties (iii) and (iv), the flow map Φ_t has additional less obvious structural properties which follow from the fact that the differential equation (1.2) can be represented in a Hamiltonian form, i.e.,

$$\begin{pmatrix} \dot{\boldsymbol{q}} \\ \dot{\boldsymbol{p}} \end{pmatrix} = \boldsymbol{J} \nabla H(\boldsymbol{q}, \boldsymbol{p}), \tag{1.3}$$

where $\boldsymbol{J} \in \mathbb{R}^{2n \times 2n}$ is a skew-symmetric matrix, called the symplectic structure matrix, which in the case of (1.2), takes the canonical form

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_n \\ -\boldsymbol{I}_n & \boldsymbol{0} \end{pmatrix},$$

as the Hamiltonian H is separable in the sense that the total energy decomposes as the sum of kinetic energy and potential energy. A fundamental property of solutions of Hamiltonian differential equations is that the collection $(\Phi_t)_{t \in \mathbb{R}}$ of associated flow maps has a symplectic group structure [73], which means that the symplectic 2-form is preserved under the action of each group element, i.e.,

$$\nabla \Phi_t{}^T \boldsymbol{J} \nabla \Phi_t = \boldsymbol{J}, \text{ for all } t \in \mathbb{R}. \tag{1.4}$$

where $\nabla \Phi_t$ denotes the Jacobian of Φ_t. From (1.4) it follows that $|\det(\Phi_t')| = 1$. Therefore, a consequence of (1.4) is that Φ_t is volume conserving, i.e.,

$$\int_{\Omega_{\boldsymbol{x}}} \mathbf{1}_{\Phi_t(S)} \mathrm{d}x = \int_{\Omega_{\boldsymbol{x}}} \mathbf{1}_S \mathrm{d}x,$$

for all (measurable) subsets $S \subset \Omega_{\boldsymbol{x}}$.

Symplectic integration

The differential equation (1.2) can only be solved exactly in special cases, for example if $\nabla_{\boldsymbol{q}} U(\boldsymbol{q})$ is linear. Therefore, in practice trajectories are computed using a numerical integration method:

$$(\hat{\boldsymbol{q}}_{n+1}, \hat{\boldsymbol{p}}_{n+1}) = \hat{\Phi}_{\Delta t}(\hat{\boldsymbol{q}}_n, \hat{\boldsymbol{p}}_n).$$

with

$$(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{p}}_0) = (\boldsymbol{q}(0), \boldsymbol{p}(0)).$$

The dynamics of (1.2) is in general chaotic, therefore it is a computationally infeasible task to approximate the exact value of $(\boldsymbol{q}(t), \boldsymbol{p}(t))$ by the numerical approximation $(\hat{\boldsymbol{q}}(n\Delta t), \hat{\boldsymbol{p}}(n\Delta t)), t = n\Delta t$, for large time t, due to the exponential growth of the global error. However, for the computation of macroscopic averages this error is of secondary importance. Intuitively speaking, what is more important is that the geometric structure of the manifold $(\boldsymbol{q}(t), \boldsymbol{p}(t))_{t>0}$ corresponding to the trajectory of the exact solution is replicated by the the discrete approximation $(\hat{\boldsymbol{q}}_n, \hat{\boldsymbol{p}}_n)_{n \in \mathbb{N}}$. One of the most effective

ways to construct such structure preserving integration schemes is by designing the integration map $\hat{\Phi}_{\Delta t}$ in such a way that the symplectic 2-form is preserved under the discrete dynamics of the integration scheme, that is

$$\hat{\Phi}'_{\Delta t}{}^{T} \boldsymbol{J} \hat{\Phi}'_{\Delta t} = \boldsymbol{J}, \tag{1.5}$$

where $\hat{\Phi}'_{\Delta t}$ denotes the Jacobian of $\hat{\Phi}_{\Delta t}$. The reason why symplectic integration schemes do particularly well in in long term simulation is best explained by adopting a backward error analysis perspective: Let $\hat{\Phi}_{\Delta t}$ be the exact flow map of the perturbed ODE:

$$\begin{pmatrix} \dot{\boldsymbol{q}} \\ \dot{\boldsymbol{p}} \end{pmatrix} = \boldsymbol{J} \nabla H(\boldsymbol{q}, \boldsymbol{p}) + \sum_{i=1}^{\infty} \Delta t^i f_i(\boldsymbol{q}, \boldsymbol{p}).$$

If $\hat{\Phi}_{\Delta t}$ satisfies (1.5), then f_i can be shown to have a Hamiltonian structure (See e.g. [73]), i.e., $f_i = \nabla \boldsymbol{J} H_i$, so that formally $\hat{\Phi}_{\Delta t}$ can be interpreted as the exact flow of a Hamiltonian system of the form

$$H_{\Delta t} = H + \sum_{i=1}^{\infty} \Delta t^i H_i. \tag{1.6}$$

The series (1.6) in general does not converge, i.e., the remainder term of the truncated series

$$H_{\Delta t}^{(k)} = H + \sum_{i=1}^{k} \Delta t^i H_i.$$

cannot be bounded. However, under the condition that H and $H_{\Delta t}^{(k)}$ are smooth, it can however be shown [119, 13, 43] that the difference between the flow $\hat{\Phi}_{\Delta t}$ and the flow $\Phi_{\Delta t}^{(k)}$ of the Hamiltonian $H_{\Delta t}^{(k)}$ decreases exponentially, i.e., for all $(\boldsymbol{q}, \boldsymbol{p}) \in \Omega_{\boldsymbol{q}} \times \mathbb{R}^n$

$$\|\hat{\Phi}_{\Delta t}(\boldsymbol{q}, \boldsymbol{p}) - \Phi_{\Delta t}^{(k)}(\boldsymbol{q}, \boldsymbol{p})\|_2 = O(e^{-\frac{1}{\Delta t}}). \tag{1.7}$$

For sufficiently small Δt and appropriately chosen k (See e.g. [73, 44] for more details on how to choose the truncation order k in an optimal way), this residual error is so small that the resulting systematic energy drift is not noticeable on the time-scale of most molecular simulations [36]. Moreover, if H and $H_{\Delta t}^{(k)}$ are Lipschitz, it is easy to see that for any $E_0 = H(\boldsymbol{q}(0), \boldsymbol{p}(0))$, the distance between the manifolds

$$\mathcal{M}_{H,E_0} = \{(\boldsymbol{q}, \boldsymbol{p}) \in \Omega_{\boldsymbol{q}} \times \mathbb{R}^n : H(\boldsymbol{q}, \boldsymbol{p}) = E_0\}, \tag{1.8}$$

and

$$\mathcal{M}_{H_{\Delta t}^{(k)},E_0} = \{(\boldsymbol{q}, \boldsymbol{p}) \in \Omega_{\boldsymbol{q}} \times \mathbb{R}^n : H_{\Delta t}^{(k)}(\boldsymbol{q}, \boldsymbol{p}) = E_0\},$$

can be uniformly controlled by Δt, which explains the structure preserving property of symplectic integrators. Finally, this also implies that up to the small energy drift caused by the residual error (1.7), fluctuations in $H(\hat{\boldsymbol{q}}_n, \hat{\boldsymbol{q}}_n) - E_0, n \in \mathbb{N}$ are bounded and likewise can be uniformly controlled by Δt.

One of the most commonly used symplectic numerical integrators is the Störmer-

Verlet method [122], which in algorithmic form reads

$$\begin{aligned} \boldsymbol{q}_{k+1/2} &= \boldsymbol{q}_k + \frac{\Delta t}{2}\boldsymbol{M}^{-1}\boldsymbol{p}_k, \\ \boldsymbol{p}_{k+1} &= \boldsymbol{p}_k - \Delta t \nabla U(\boldsymbol{q}_{k+1/2}), \\ \boldsymbol{q}_{k+1} &= \boldsymbol{q}_{k+1/2} + \frac{\Delta t}{2}\boldsymbol{M}^{-1}\boldsymbol{p}_{k+1}. \end{aligned} \tag{1.9}$$

The Störmer-Verlet method is constructed as a symmetric splitting scheme [44], i.e., it is constructed from a decomposition of the Liouville operator $\mathcal{L}_H^\dagger$ associated with the Hamiltonian vector field $\boldsymbol{J}\nabla H(\boldsymbol{q},\boldsymbol{p})$ as

$$\mathcal{L}_{\mathrm{H}}^\dagger = \mathcal{L}_{\mathrm{A}}^\dagger + \mathcal{L}_{\mathrm{B}}^\dagger,$$

where

$$\mathcal{L}_{\mathrm{A}}^\dagger = \nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \cdot \nabla_{\boldsymbol{p}}, \quad \mathcal{L}_{\mathrm{B}}^\dagger = -\boldsymbol{M}^{-1}\boldsymbol{p} \cdot \nabla_{\boldsymbol{q}}.$$

using a symmetric Strang-splitting beginning with an integration of the vector field associated with $\mathcal{L}_A$. Formally, one can therefore write the flow map $\mathscr{F}_{\Delta t}$ associated with the recursion (1.9) defined by

$$\mathscr{F}_{\Delta t}^{\mathrm{BAB}} = e^{\mathcal{L}_{\mathrm{B}}^\dagger \Delta t/2} e^{\mathcal{L}_{\mathrm{A}}^\dagger \Delta t} e^{\mathcal{L}_{\mathrm{B}}^\dagger \Delta t/2}.$$

From the construction of the Verlet method as a symmetric splitting method one can directly deduce that its convergence order in Δt is of order two (see e.g. [73, 44] for details.) Symplectic integrators of arbitrary orders can be constructed as splitting schemes via combination techniques due to Suzuki [117] and Yoshida [127].

1.1.3 Thermodynamic statistical ensembles

The purpose of molecular dynamics simulation typically lies in the computation of macroscopic quantities. So far, we have described these loosely as averaged quantities. Within the framework of statistical physics this notion is made precise by introducing the concept of a macro-state or statistical ensemble. In mathematical terms, a macro-state is simply a probability measure μ, which is defined on the phase space $\Omega_{\boldsymbol{q}} \times \mathbb{R}^n$ and macroscopic features, such as temperature, pressure are understood as expectations/averages of observables with respect to this probability measure, i.e.,

$$\mathbb{E}_\mu[\varphi] = \int_{\Omega_{\boldsymbol{q}} \times \mathbb{R}^n} \varphi(\boldsymbol{q},\boldsymbol{p}) \mu(\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}),$$

where $\varphi \in L^1(\mu)$ denotes an observable.

1.1.4 The micro-canonical ensemble

The micro-canonical ensemble, describes an isolated system at constant energy in a fixed volume, where the number of particles does not change over time. It is therefore also termed the NVE ensemble, in reference to these three preserved quantities. Hamiltonian molecular dynamics as we have described it in the previous sections exactly corresponds to this setup. It is therefore also not surprising that the associated macro-state μ_E is a probability measure, whose mass collapses on the manifold defined by the level set of

a prescribed energy value E, i.e.,

$$\mu_E(\mathrm{d}\boldsymbol{q}, \mathrm{d}\boldsymbol{p}) = \frac{1}{Z_E}\delta(H(\mathrm{d}\boldsymbol{q}, \mathrm{d}\boldsymbol{p}) - E),$$

which is to be understood as a generalisation of a Dirac measure with $\delta_{H(\boldsymbol{q},\boldsymbol{p})-E}(\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p})$ being defined as

$$\int_{\Omega_{\boldsymbol{q}}\times\mathbb{R}^n} \varphi(\boldsymbol{q},\boldsymbol{p})\delta(H(\mathrm{d}\boldsymbol{q}, \mathrm{d}\boldsymbol{p}) - E) := \lim_{\Delta E\to 0} \frac{1}{2\Delta E}\int_{\Omega_{\boldsymbol{q}}\times\mathbb{R}^n} \mathbf{1}_{M_{E,\Delta E}}(\boldsymbol{q},\boldsymbol{p})\varphi(\boldsymbol{q},\boldsymbol{p})\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p},$$

with

$$M_{E,\Delta E} := \{(\boldsymbol{q},\boldsymbol{p}) \in \Omega_{\boldsymbol{q}} \times \mathbb{R}^n : E - \Delta E \leq H(\boldsymbol{q},\boldsymbol{p}) \leq E + \Delta E\}.$$

and normalisation constant

$$Z_E = \int_{\Omega_{\boldsymbol{q}}\times\mathbb{R}^n} \delta(H(\mathrm{d}\boldsymbol{q}, \mathrm{d}\boldsymbol{p}) - E).$$

1.1.5 Canonical ensemble

Laboratory conditions which are to be replicated with molecular dynamics simulations are typically such that the system under study is not isolated, but is surrounded by other matter, with which it exchanges energy. As a consequence the total energy fluctuates over time, but temperature becomes a well defined macroscopic quantity [121]. If the temperature is assumed to be constant in time the system is said to be in equilibrium and the micro-states of the system, i.e., the position $\boldsymbol{q}$ and momentum $\boldsymbol{p}$ of the particles are distributed according to the canonical distribution, the density for which takes the form

$$\mu_\beta(\mathrm{d}\boldsymbol{q}, \mathrm{d}\boldsymbol{p}) = \frac{1}{Z}e^{-\beta H(\boldsymbol{q},\boldsymbol{p})}\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p}, \tag{1.10}$$

where β is the reciprocal of the kinetic temperature T_s scaled by Boltzmann's constant k_B, i.e.

$$\beta^{-1} = k_\mathrm{B}T_s := n^{-1}\mathbb{E}_{\mu_\beta}[\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p}]. \tag{1.11}$$

The normalisation constant Z, also referred to the partition function, is

$$Z = \int_{\Omega_{\boldsymbol{q}}\times\mathbb{R}^n} e^{-\beta H(\boldsymbol{q},\boldsymbol{p})}\mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p}.$$

The partition function Z is in general not known. The computation of Z is considered as a hard problem. A comprehensive introduction to thermodynamic integration can be found in e.g. [75, Chaptetr 2]. The canonical ensemble is also sometimes referred to as the NVT ensemble since it is defined such that the number N of particles, the volume V, and the temperature T_s, are all constant quantities. A derivation of the form (1.10) of the canonical ensemble is beyond the scope of this introduction. Let us however summarise a few properties:

(i) The canonical distribution maximises entropy. More precisely, let $\rho \in L^1(\Omega_{\boldsymbol{q}} \times \mathbb{R}^n)$. It can be shown (See e.g. [75]), that for a given energy value E, the optimisation problem

$$\rho_{\max} := \operatorname*{arg\,max}_{\rho\in L^1(\Omega_{\boldsymbol{q}}\times\mathbb{R}^n)} \mathcal{S}(\rho),$$

where ρ is subject to the constraints

$$\rho \geq 0, \quad \int_{\Omega_{\boldsymbol{q}} \times \mathbb{R}^n} \rho(\boldsymbol{q},\boldsymbol{p}) \mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p} = 1, \quad \int_{\Omega_{\boldsymbol{q}} \times \mathbb{R}^n} H(\boldsymbol{q},\boldsymbol{p})\rho(\boldsymbol{q},\boldsymbol{p}) \mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p} = E, \tag{1.12}$$

and $\mathcal{S}$ denotes the Gibb-Boltzmann entropy, i.e.,

$$\mathcal{S}(\rho) = -\int_{\Omega_{\boldsymbol{q}} \times \mathbb{R}^n} \rho(\boldsymbol{q},\boldsymbol{p}) \ln \rho(\boldsymbol{q},\boldsymbol{p}) \mathrm{d}\boldsymbol{q}\mathrm{d}\boldsymbol{p},$$

has the density of the Gibb-Boltzmann distribution (1.10) as its unique solution, i.e.,

$$\rho_{\max} = \frac{1}{Z} e^{-\beta H(\boldsymbol{q},\boldsymbol{p})} =: \rho_\beta(\boldsymbol{q},\boldsymbol{p}).$$

(ii) The above interpretation of β as the inverse temperature is consistent with the more elementary definition of the temperature T_s as

$$\frac{1}{k_{\mathrm{B}} T_s} = \frac{\partial \mathcal{S}}{\partial E}.$$

The inverse temperature β can be identified with a wide range of observables as the following proposition shows (See [57, 67], also [71] for a proof)

Proposition 1.1.1. *Let $G \in C^1(\Omega_{\boldsymbol{q}} \times \mathbb{R}^n, \mathbb{R}^n \times \mathbb{R}^n)$ be a vector field with the following properties*

(i) $0 < \left|\mathbb{E}_{\mu_\beta}[G(\boldsymbol{q},\boldsymbol{p}) \cdot \nabla H(\boldsymbol{q},\boldsymbol{p})]\right| < \infty$,

(ii) $0 < \mathbb{E}_{\mu_\beta}[\nabla \cdot G(\boldsymbol{q},\boldsymbol{p})] < \infty$,

(iii) $\sup_{(\boldsymbol{q},\boldsymbol{p}) \in \mathbb{R}^{2n}} \left\| G(\boldsymbol{q},\boldsymbol{p}) e^{-\beta H(\boldsymbol{q},\boldsymbol{p})} \right\|_2 < \infty$,

then

$$\beta^{-1} = \frac{\mathbb{E}_{\mu_\beta}[G(\boldsymbol{q},\boldsymbol{p}) \cdot \nabla H(\boldsymbol{q},\boldsymbol{p})]}{\mathbb{E}_{\mu_\beta}[\nabla \cdot G(\boldsymbol{q},\boldsymbol{p})]}$$

The choice $G(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{0},\boldsymbol{p})$ gives the identity (1.11). Other valid choices are for example $G(\boldsymbol{q},\boldsymbol{p}) = (\boldsymbol{q}_i \boldsymbol{e}_i, \boldsymbol{0})$, which implies

$$\beta^{-1} = \mathbb{E}_{\mu_\beta}[\boldsymbol{q}_i \partial_{\boldsymbol{q}_i} U(\boldsymbol{q})], \tag{1.13}$$

and hence in particular

$$\beta^{-1} = n^{-1} \mathbb{E}_{\mu_\beta}[\boldsymbol{q} \cdot \nabla U(\boldsymbol{q})],$$

which is also known as the Virial theorem [42]. Similarly, $G(\boldsymbol{q},\boldsymbol{p}) = (\nabla U(\boldsymbol{q}), \boldsymbol{0})$ satisfies the condition of Proposition 1.1.1 yielding the identity

$$\beta^{-1} = \frac{\mathbb{E}_{\mu_\beta}\left[\|\nabla U(\boldsymbol{q})\|_2^2\right]}{\mathbb{E}_{\mu_\beta}\left[\Delta U(\boldsymbol{q})\right]} \tag{1.14}$$

which corresponds to Rugh's formulation of configurational temperature in [107]. One can generalise Proposition 1.1.1 to cover also the case of periodic boundary conditions, i.e., $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n$. In this case ones must additionally require that the terms $G \cdot \nabla H$ and $\nabla \cdot G$ are translation invariant in $\boldsymbol{q}$ (See [71]).

1.2 Bayesian inference

Let $X_{1:N}$ be a sequence of random variables taking values in a measurable space $(\mathbb{X}, \mathcal{X})$, independently and identically distributed with density $p(\cdot|\boldsymbol{\theta})$ with respect to some (fixed) dominating measure on $\mathbb{X}$, where $\boldsymbol{\theta}$ is a parameter in $\mathbb{R}^n$. Given a prior for $\boldsymbol{\theta}$ with density π and observed values $x_{1:N}$ of the random variables $X_{1:N}$, the posterior distribution over $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta}|x_{1:N}) = \frac{\pi(\boldsymbol{\theta}) \prod_{i=1}^N p(x_i|\boldsymbol{\theta})}{\int_{\mathbb{R}^n} \pi(\boldsymbol{\theta}) \prod_{i=1}^N p(x_i|\boldsymbol{\theta}) d\boldsymbol{\theta}}. \tag{1.15}$$

In the above expression $p(x_i|\boldsymbol{\theta})$ if interpreted as a function of θ is referred to as the *likelihood function* defined by a *statistical model*. Statistical inference which is based on the properties of the posterior distribution $p(\boldsymbol{\theta}|x_{1:N})$ is referred to as Bayesian inference. Similarly as in the calculation of macroscopic feature in the context of molecular dynamics, one is typically interested in the calculation of expectations of the form

$$\mathbb{E}_{p(\boldsymbol{\theta}|x_{1:N})}\left[\varphi(\boldsymbol{\theta})\right] = \int_{\mathbb{R}^n} \varphi(\boldsymbol{\theta}) p(\boldsymbol{\theta}|x_{1:N}) \mathrm{d}\boldsymbol{\theta} \tag{1.16}$$

for $\varphi \in L^1(p(\cdot|x_{1:N}))$. For example, one might be interested in the calculation of credibility regions around a estimate $\widehat{\boldsymbol{\theta}}$, which would require the calculation of probabilities of the from

$$\mathbb{P}_{p(\boldsymbol{\theta}|x_{1:N})}(\boldsymbol{\theta} \in A) = \mathbb{E}_{p(\boldsymbol{\theta}|x_{1:N})}\left[\mathbf{1}_A(\boldsymbol{\theta})\right]$$

for $A \in \mathcal{B}(\mathbb{R}^n)$ with $\widehat{\boldsymbol{\theta}} \in A$. A very common task in Bayesian inference is classification. In this situation the observations $x_{1:N}$ come in the form of paired data, that is $x_i = (y_i, z_i)$, which are associated with the random variables Y_i and Z_i, respectively. Typically, Y_i is referred to as a predictor variable and Z_i as a response variable, whose observations $z_i \in \mathcal{C} \subset \mathbb{Z}$ are understood as class labels. The likelihood function, $p(x_i|\boldsymbol{\theta})$, in the expression of the posterior in (1.15) then takes the form

$$p(x_i \mid \boldsymbol{\theta}) = p(z_i \mid y_i, \boldsymbol{\theta}),$$

where the expression $p(z_i \mid y_i, \boldsymbol{\theta})$ is specified by a statistical model. The statistical model is parametrised by $\boldsymbol{\theta}$ and describes in a probabilistic sense the relation between Y_i and Z_i. A prediction of a new observation $(\tilde{y}, \tilde{z})$ associated with the pair of random variables $(\tilde{Y}, \tilde{Z})$, is then computed by the expectation of the likelihood $p(\tilde{z} \mid \tilde{y}, \boldsymbol{\theta})$ under the posterior density $p(\boldsymbol{\theta}|x_{1:N})$, i.e.,

$$\mathbb{P}(\tilde{Z} = \tilde{z} \mid \tilde{Y} = \tilde{y}, \{y_i, z_i\}_{i=1}^N) = \mathbb{E}_{p(\boldsymbol{\theta}|x_{1:N})}[p(\tilde{z} \mid \tilde{y}, \boldsymbol{\theta})].$$

The choice and specification of the statistical model is a modelling decision. Often these models are described by a graphical model, such as a Bayesian network or a Gaussian random field (See e.g. [16, 91, 9] for a general introduction) and the associated likelihood function can for example be described by a neural network (See e.g. [92]), a logistic regression function (See Section 1.2.2), or a mixture of Gaussian densities (See Section 1.2.1).

1.2.1 Bayesian Gaussian mixture model

Consider paired data of the form $(x_i, z_i)_{i=1}^N$ where $x_i \in \mathbb{R}^{n_y}$ and class labels $z_i \in \mathcal{C} = \{0, 1, \ldots, N_c\} \subset \mathbb{N}$. A very common way to model such data is to assume that X_i is distributed according to a Gaussian mixture distribution. More precisely, it is assumed that

1. the class labels are distributed according to a multivariate Bernoulli distribution, i.e.,
$$\mathbb{P}(Z_i = k) = \boldsymbol{w}_k, k \in \mathcal{C}, \tag{1.17}$$
with $\boldsymbol{w}_k > 0, k = 1, \ldots, N_c$ and $\sum_{k=1}^{N_c} \boldsymbol{w}_k = 1$.

2. the distribution of the conditional probabilities $\mathbb{P}(X_i | Z_i = k), k \in \mathcal{C}$ are Gaussian, i.e.,
$$X_i \mid Z_i = k \ \sim \ \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$
where $\mu_k \in \mathbb{R}^{n_y}$ and $\boldsymbol{\Sigma}_k \in \mathbb{R}^{n_y \times n_y}$ symmetric positive definite, denote the mean and covariance of a multivariate Gaussian density.

This statistical model corresponds to a generative model where the data is generated in two steps: In the first step the class label z_i is sampled according to the multivariate Bernoulli distribution (1.17). In the second step the response x_i is sampled from the multivariate Gaussian distribution corresponding to the class label z_i.
Given the parameterisation $\boldsymbol{\theta} = (\boldsymbol{w}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_{1 \le k \le N_c}$ of the model, it follows from Bayes' Theorem that the conditional probability of the response Z_i to take the value of the observation z_i given the value x_i of the predictor variable is

$$\mathbb{P}\left(Z_i = z_i \mid X_i = x_i, \ (\boldsymbol{w}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_{1 \le k \le N_c}\right) = \frac{\boldsymbol{w}_{z_i} \mathcal{N}(x_i; \boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i})}{\sum_{k=1}^{N_c} \boldsymbol{w}_k \mathcal{N}(x_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}.$$

In a Bayesian setup the model parameters $\boldsymbol{w}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ are treated as random variables. Assuming a prior distribution with density π of the generic form

$$\pi = \prod_{k=1}^{N_c} \pi_k, \ (\boldsymbol{w}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \sim \pi_k,$$

the corresponding posterior density for the observations $(y_i, z_i)_{1 \le i \le N}$ follows again from Bayes' Theorem:

$$p((\boldsymbol{w}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)_{1 \le k \le N_c} | (y_i, z_i)_{1 \le i \le N}) \propto \prod_{k=1}^{N_c} \pi_k \left(\boldsymbol{w}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \prod_{i=1}^{N} \left(\frac{\boldsymbol{w}_{z_i} \mathcal{N}(y_i; \boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i})}{\sum_{k=1}^{N_c} \boldsymbol{w}_k \mathcal{N}(y_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \right)$$

1.2.2 Bayesian logistic regression

Consider a binary classification problem, for a given set of paired data $(y_i, z_i)_{1 \le i \le N}$, where $y_i \in \mathbb{R}^n$ denotes the predictor variable and $z_i \in \{-1, 1\}$ the response variable. Let X_i and Z_i as described above be i.i.d. random variables associated with the observations (y_i, z_i), respectively. A common way to model the dependence between the predictor variable and the response variable in binary classification problems is to assume that the conditional probability of the event $Z_i = 1$ given X_i is described by the values of the logistic function evaluated at a linear transformation of X_i, i.e.,

$$\mathbb{P}(Z_i = 1 \mid X_i) = f(\boldsymbol{w}^T X_i),$$

hence

$$\mathbb{P}(Z_i \mid X_i) = f(Z_i \boldsymbol{w}^T X_i),$$

where

$$f(s) = \frac{1}{1 + e^{-s}},$$

is the logistic function and the weights $\boldsymbol{w} \in \mathbb{R}^n$ define the linear transformation. In a Bayesian setup, the weights are considered to be random variables distributed according to a prior distribution, $\boldsymbol{w} \sim \pi$. By Bayes rule, the posterior distribution over $\boldsymbol{w}$ is then found to be,

$$p(\boldsymbol{w} \mid x, y) \propto \pi(\boldsymbol{w}) p(y \mid x, \boldsymbol{w}) = \pi(\boldsymbol{w}) \prod_{i=1}^{N} f(z_i \boldsymbol{w}^T y_i).$$

and corresponding log posterior function is simply

$$\log p(\boldsymbol{w} \mid x, y) = \log \pi(\boldsymbol{w}) \sum_{i=1}^{N} \log f(z_i \boldsymbol{w}^T y_i) + const.$$

We consider a Bayesian logistic regression model in Section 5.3.2, where we use it for inference on the MNIST data set [68].

1.2.3 Sampling in Bayesian statistics

For most statistical models expectations against the posterior distribution of the form (1.16) don't take a closed form and the computation of these via numeral quadrature is computationally infeasible due to the typically very high dimension n of the parameter space. Instead, it is only possible to evaluate the expression

$$p(x_{1:N} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}),$$

which is proportional to the posterior density, pointwise. Therefore, as for the computation of macroscopic features in molecular dynamics, Monte Carlo methods are required in order to approximate the respective expectation by sampling from the posterior density $p(\boldsymbol{\theta} \mid x_{1:N})$. Sampling in Bayesian inference applications is conceptually identical to the sampling of macroscopic features corresponding to stationary averages in molecular dynamics. More precisely, as long as the unnormalised posterior density is well defined,[1] the negative logarithm of the unnormalised posterior defines a confining potential function, i.e.,

$$U(\boldsymbol{\theta}) = -\log \pi(\boldsymbol{\theta}) - \sum_{i=1}^{N} \log p(x_{1:N} \mid \boldsymbol{\theta}). \tag{1.18}$$

In the remainder of this section we discuss the construction of Markov Chain Monte-Carlo (MCMC) methods via the Metropolis-Hastings algorithm. Although the presentation is given in reference to the Bayesian sampling application, the concepts can be applied similarly in the molecular dynamics context.

[1] In the sense that the expression it is integrable with respect to the Lebesque measure

1.2.4 The Metropolis-Hastings algorithm

Classical MCMC algorithms produce samples approximately distributed according to a given measure, which then allow estimators of the quantities of interest mentioned above to be produced. The Metropolis-Hastings (MH) algorithm is a common means of constructing a transition kernel giving rise to such a Markov chain; a transition density $\rho_{trans} : \mathbb{R}^n \times \mathbb{R}^n \to [0, 1]$ is supplied, and a new transition kernel $P : \mathbb{R}^n \times \mathcal{B}(\mathbb{R}^n) \to [0, 1]$ is produced according to the following relationship:

$$P(\boldsymbol{\theta}, \mathrm{d}\boldsymbol{\theta}') = \left[\int_{\mathbb{R}^n} (1 - p_{acc}(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta}))\rho_{trans}(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta})\mathrm{d}\widetilde{\boldsymbol{\theta}}\right] \delta_{\boldsymbol{\theta}}(\mathrm{d}\boldsymbol{\theta}') + p_{acc}(\boldsymbol{\theta}'|\boldsymbol{\theta})\rho_{trans}(\boldsymbol{\theta}'|\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}'$$

where the acceptance ratio p_{acc} is given by

$$p_{acc}(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta}) = \min\left(1, \frac{p(\widetilde{\boldsymbol{\theta}}|x_{1:N})\rho_{trans}(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}})}{p(\boldsymbol{\theta}|x_{1:N})\rho_{trans}(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta})}\right) \tag{1.19}$$

Provided that (1.19) is well defined (i.e.,in particular the support of

$$p(\widetilde{\boldsymbol{\theta}}|x_{1:N})\rho_{trans}(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}})$$

and

$$p(\boldsymbol{\theta}|x_{1:N})\rho_{trans}(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$$

must coincide), then under certain conditions (see e.g., [86] and Section 2.6), the Markov chain associated with P ergodically samples the invariant distribution $p(\boldsymbol{\theta}|x_{1:N})$. The accept-reject step in the Metropolis-Hastings algorithm ensures that the generated Markov chain satisfies detailed balance with respect to the target density $p(\boldsymbol{\theta}|x_{1:N})$, i.e.,

$$P(\boldsymbol{\theta}, \mathrm{d}\boldsymbol{\theta}')p(\boldsymbol{\theta}|x_{1:N})\mathrm{d}\boldsymbol{\theta} = P(\boldsymbol{\theta}', \mathrm{d}\boldsymbol{\theta})p(\boldsymbol{\theta}'|x_{1:N})\mathrm{d}\boldsymbol{\theta}'. \tag{1.20}$$

The structure of the MH algorithm is given in Algorithm 1.

Algorithm 1 The Metropolis-Hastings algorithm

Require: T, ρ_{trans}, $\boldsymbol{\theta}_0$, $p(\cdot|x_{1:N})$
  for $t = 1, \ldots, T$ **do**
    Draw $\widetilde{\boldsymbol{\theta}}_t \sim \rho_{trans}(\cdot|\boldsymbol{\theta}_{t-1})$
    Set $\boldsymbol{\theta}_t \coloneqq \widetilde{\boldsymbol{\theta}}_t$ w.p. $p_{acc}(\widetilde{\boldsymbol{\theta}}_t|\boldsymbol{\theta}_{t-1})$, else set $\boldsymbol{\theta}_t \coloneqq \boldsymbol{\theta}_{t-1}$
  end for
  Return $(\boldsymbol{\theta}_t)_{t=0}^T$

Common choices for the proposal distribution $\rho_{trans}(\boldsymbol{\theta}'|\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}'$ include

$$\boldsymbol{\theta}' = \boldsymbol{\theta} + \sqrt{\Delta t}\mathscr{R} \tag{1.21}$$

$$\boldsymbol{\theta}' = \boldsymbol{\theta} - \Delta t \nabla U(\boldsymbol{\theta}) + \sqrt{2\Delta t}\mathscr{R} \tag{1.22}$$

where $\Delta t > 0$ and $\mathscr{R}$ is some mean-zero random variable with unit covariance matrix $\mathbf{I}_n$. The former leads to the random walk Metropolis-Hastings algorithm (RWMH), whilst the latter leads to the Metropolis-adjusted Langevin algorithm (MALA). A large number of variants of these algorithms have been proposed and studied in the literature, many of which include the addition of an auxiliary variable which in the statistical physics corresponds to the momentum variable $\boldsymbol{p}$. For example in Hamiltonian Monte

Carlo (HMC) [30] the proposal is constructed using the approximate Hamiltonian flow of the Verlet scheme (1.9). The Markov chain associated with the MALA kernel can be viewed as the Euler-Maruyama discretization applied to the following (overdamped) Langevin stochastic differential equation

$$\dot{\boldsymbol{\theta}} = -\nabla U(\boldsymbol{\theta}) + \sqrt{2}\boldsymbol{\eta}(t) \tag{1.23}$$

with an acceptance step added so that the invariant distribution is not affected by the discretization. This idea can be generalised and a large variety of Metropolis-Hastings algorithm can be constructed using discretisations of ergodic SDEs as proposals.

1.2.5 The generalised Metropolis-Hastings algorithm

In situations where a discretisation of an ergodic SDE with a non-reversible solution process is used as a proposal, imposing detailed balance via a standard Metropolis acceptance-rejection criteria can result in very poor acceptance rates as the ratio of the proposal densities will typically be very small. In this case the accept-reject step can be adjusted such that a modified detailed balance condition is satisfied. Variants of the Metropolis-Hastings algorithm which preserve a modified detailed balance condition are referred to as generalised Metropolis-Hastings algorithms. In what follows we describe the basic construction of such algorithms. For a detailed discussion we refer to [75, Chapter 2].

Let $\boldsymbol{x} := (\boldsymbol{\theta}, \boldsymbol{s})$ with $\boldsymbol{s} \in \mathbb{R}^m$ with $m \in \mathbb{N}$ and $\rho(\boldsymbol{x}) = \rho(\boldsymbol{\theta}, \boldsymbol{s}) = p(\boldsymbol{\theta}|x_{1:N})\nu(\boldsymbol{s})$, a probability density which is the product of the target density $p(\boldsymbol{\theta}|x_{1:N})$ and some smooth probability density $\nu \in \mathcal{C}^\infty(\mathbb{R}^m, \mathbb{R})$, so that

$$p(\boldsymbol{\theta}|x_{1:N}) = \int_{\mathbb{R}^n} \rho(\boldsymbol{\theta}, \boldsymbol{s})\mathrm{d}\boldsymbol{s}.$$

Let further $S : \mathbb{R}^m \to \mathbb{R}^m$ be an involutive transformation, i.e. $S = S^{-1}$, which in addition leaves the measure associated with ρ invariant, i.e.,

$$\rho(S(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} = \rho(S^{-1}(\boldsymbol{x}))\mathrm{d}\boldsymbol{x} = \nu(\boldsymbol{x})\mathrm{d}\boldsymbol{x}.$$

For example, if $\nu(\boldsymbol{s}) \propto \exp(-\boldsymbol{s}\cdot\boldsymbol{s})$, then the involutive transformation $S(\boldsymbol{\theta}, \boldsymbol{s}) = (\boldsymbol{\theta}, -\boldsymbol{s})$, satisfies this property. In a generalised Metropolis-Hastings algorithm the transition kernel $P : \mathbb{R}^{n+m} \times \mathcal{B}(\mathbb{R}^{n+m}) \to [0, 1]$ is constructed in such a way that a modified detailed balance condition of the form

$$P(\boldsymbol{x}, \mathrm{d}\boldsymbol{x}')\rho(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = P(S(\boldsymbol{x}'), S(\mathrm{d}\boldsymbol{x}))\rho(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}'. \tag{1.24}$$

Note that the modified detailed balance condition (1.24) indeed implies that the measure $\rho(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ is preserved under the Markov chain associated withe the kernel $P(\boldsymbol{x}, \mathrm{d}\boldsymbol{x}')$.

This follows since for any smooth test function φ with compact support one finds

$$\begin{aligned}\int_{\mathbb{R}^{n+m}}\int_{\mathbb{R}^{n+m}}\varphi(\boldsymbol{x}')P(\boldsymbol{x},\mathrm{d}\boldsymbol{x}')\rho(\boldsymbol{x})\mathrm{d}\boldsymbol{x} &= \int_{\mathbb{R}^{n+m}}\int_{\mathbb{R}^{n+m}}\varphi(\boldsymbol{x}')P(S(\boldsymbol{x}'),S(\mathrm{d}\boldsymbol{x}))\rho(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' \\ &= \int_{\mathbb{R}^{n+m}}\left(\int_{\mathbb{R}^{n+m}}P(S(\boldsymbol{x}'),S(\mathrm{d}\boldsymbol{x}))\right)\varphi(\boldsymbol{x}')\rho(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' \\ &= \int_{\mathbb{R}^{n+m}}\varphi(\boldsymbol{x}')\rho(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}'.\end{aligned} \tag{1.25}$$

A simple calculation shows that the modified detailed balance condition (1.24) is satisfied for a memory kernel of the form

$$P(\boldsymbol{x},\mathrm{d}\boldsymbol{x}') = \left[\int_{\mathbb{R}^n}(1-p_{acc}(\widetilde{\boldsymbol{x}}|\boldsymbol{x}))\rho_{trans}(\widetilde{\boldsymbol{x}}|\boldsymbol{x})\mathrm{d}\widetilde{\boldsymbol{x}}\right]\delta_{S(\boldsymbol{x})}(\mathrm{d}\boldsymbol{x}') + p_{acc}(\boldsymbol{x}'|\boldsymbol{x})\rho_{trans}(\boldsymbol{x}'|\boldsymbol{x})\mathrm{d}\boldsymbol{x}' \tag{1.26}$$

where

$$p_{acc}(\widetilde{\boldsymbol{x}}|\boldsymbol{x}) = \min\left(1,\frac{\rho(\widetilde{\boldsymbol{x}})\rho_{trans}(S(\boldsymbol{x})|S(\widetilde{\boldsymbol{x}}))}{\rho(\boldsymbol{x})\rho_{trans}(\widetilde{\boldsymbol{x}}|\boldsymbol{x})}\right). \tag{1.27}$$

Thus, the Markov chain generated by the following Algorithm 2 preserves the measure $\rho(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$. A well known application of the generalised Metropolis-Hastings algorithm is

Algorithm 2 The generalised Metropolis-Hastings algorithm

Require: T, S, ρ_{trans}, $\boldsymbol{x}_0$, $p(\cdot|x_{1:N})$
  for $t = 1,\ldots,T$ **do**
    Draw $\widetilde{\boldsymbol{x}}_t \sim \rho_{trans}(\cdot|\boldsymbol{x}_{t-1})$
    Set $\boldsymbol{x}_t \coloneqq \widetilde{\boldsymbol{x}}_t$ w.p. $p_{acc}(\widetilde{\boldsymbol{x}}_t|\boldsymbol{x}_{t-1})$, else set $S(\boldsymbol{x}_t) \coloneqq \boldsymbol{x}_{t-1}$
  end for
  Return $(\boldsymbol{x}_t)_{t=0}^T$

the Metropolisation of numerical schemes of the underdamped Langevin equation (See e.g. [17, 5]). In the view of the property (ii) (stated in Section 1.1.2) of the Hamiltonian flow, the transformation $S(\boldsymbol{\theta},\boldsymbol{p}) = (\boldsymbol{\theta},-\boldsymbol{p})$, is a natural choice in this case. In particular in the situation where the friction coefficient in the underdamped Langevin equation is small one can expect a scheme which is metropolised using algorithm 2 with S corresponding to a momentum flipto lead to far higher acceptance rates of proposals in comparison to the acceptance rates of proposals of a scheme which is obtained using the standard Metropolis-Hastings algorithm algorithm 1.

1.3 The generalized Langevin equation

Consider the situation of an open system exchanging energy with a heat bath. If there is no strong time scale separation between the dynamics of the heat bath and the explicitly modelled degrees of freedom, the exchange of energy between these two systems is not well modelled by a Markovian process, i.e., dynamic observables such as transport coefficients, first passage times can not be expected to be well reproduced by a thermostat model which relies on a Markovian approximation of the heat bath. For example, if we consider a distinguished particle surrounded by solvent particles of approximately the same mass, a reduced model where the interaction between the distinguished particle and the solvent particles is substituted by an underdamped Langevin equation would lead to a poor approximation of the dynamics of the distinguished particle on

the relevant time scales. In such modelling situations it is necessary to explicitly incorporate memory effects, i.e., non-markovian random forces and history dependent dissipation. The framework in which such models are typically formulated is the generalised Langevin. In this thesis we consider two different types of generalised Langevin equations, which both can be considered as non-Markovian thermostat models.

We first consider a generalised Langevin equation of the generic form

$$\begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{M}^{-1}\boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) - \int_0^t \boldsymbol{K}(t-s)\boldsymbol{M}^{-1}\boldsymbol{p}(s)\mathrm{d}s + \boldsymbol{\eta}(t). \end{aligned} \tag{1.28}$$

where the dynamic variables $\boldsymbol{q} \in \Omega_{\boldsymbol{q}}, \boldsymbol{p} \in \Omega_{\boldsymbol{p}} \subset \mathbb{R}^n$ denote the position and momentum of a Hamiltonian system with energy function

$$H(\boldsymbol{q}, \boldsymbol{p}) = U(\boldsymbol{q}) + \frac{1}{2}\boldsymbol{p}^T \boldsymbol{M}^{-1}\boldsymbol{p}, \tag{1.29}$$

where the mass tensor $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ is required to be symmetric positive definite and $U \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}}, \mathbb{R})$ is a smooth potential function. $\boldsymbol{K} : [0, \infty) \to \mathbb{R}^{n\times n}$ is a matrix valued (generalized) function, which is referred to as the memory kernel, and $\boldsymbol{\eta}$ is a stationary Gaussian process taking values in $\mathbb{R}^n$ and which (in equilibrium) is assumed to be statistically independent of $\boldsymbol{q}$ and $\boldsymbol{p}$. We refer to $\boldsymbol{\eta}$ as the noise process or random force. We further assume that a fluctuation dissipation relation between the random force $\boldsymbol{\eta}$ and the memory kernel holds so that

(i) the random force $\boldsymbol{\eta}$ is unbiased, i.e.,

$$\mathbb{E}[\boldsymbol{\eta}(t)] = \boldsymbol{0},$$

for all $t \in [0, \infty)$.

(ii) the auto-covariance function of the random force and the memory kernel $\boldsymbol{K}$ coincide up to a constant pre-factor, i.e.,

$$\mathbb{E}[\boldsymbol{\eta}(s+t)\boldsymbol{\eta}^\top(s)] = \beta^{-1}\boldsymbol{K}(t), \quad \beta > 0,$$

where the constant $\beta > 0$ corresponds to the inverse temperature of the system under consideration.

As a generalisation of the above dynamics we also consider instances of the generalised Langevin equation where the random force is a non-stationary process. More specifically, we consider the case where the strength of the random force depends on the value of the configurational variable $\boldsymbol{q}$, i.e.,

$$\begin{aligned} \dot{\boldsymbol{q}}(t) &= \boldsymbol{M}^{-1}\boldsymbol{p}(t), \\ \dot{\boldsymbol{p}}(t) &= -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}(t)) - \widetilde{\boldsymbol{K}}(\boldsymbol{q}, t) * \boldsymbol{p} + \widetilde{\boldsymbol{\eta}}(t). \end{aligned} \tag{1.30}$$

where the random force $\widetilde{\boldsymbol{\eta}}$ is assumed to be of the form

$$\widetilde{\boldsymbol{\eta}}(t) = g^T(\boldsymbol{q}(t))\boldsymbol{\eta}(t),$$

with $\boldsymbol{\eta}$ again satisfying (i) and (ii) and the convolution term, $\widetilde{\boldsymbol{K}}(\boldsymbol{q},t)*\boldsymbol{p}$, is of the form

$$\widetilde{\boldsymbol{K}}(\boldsymbol{q},t)*\boldsymbol{p} = g^T(\boldsymbol{q}(t))\int_0^t \boldsymbol{K}(t-s)g(\boldsymbol{q}(s))\boldsymbol{p}(s)\mathrm{d}s$$

with $g \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R}^{n\times n})$ and $\boldsymbol{K}$ as specified above.

The generic form of the GLE can be formally derived via a Mori-Zwanzig projection of the combined Hamiltonian dynamics of an explicit heat bath representation and the system of interest [130, 131, 88]. In what follows, we will briefly outline the Mori-Zwanzig formalism in a simplified setup following the presentation in [41]. We will then consider the particular case of the Kac-Zwanzig model and demonstrate how the above instances of the GLE can be derived from this model.

1.3.1 Formal derivation of the generalised Langevin equation via Mori-Zwanzig projection

Consider an ordinary differential equation of the form

$$\begin{aligned}\dot{\boldsymbol{u}} &= f(\boldsymbol{u},\boldsymbol{v}),\\ \dot{\boldsymbol{v}} &= g(\boldsymbol{u},\boldsymbol{v}),\end{aligned} \tag{1.31}$$

subject to the initial condition

$$(\boldsymbol{u}(0),\boldsymbol{v}(0)) = (\boldsymbol{u}_0,\boldsymbol{v}_0), \tag{1.32}$$

where f, g are smooth functions, i.e., $f \in \mathcal{C}^\infty(\mathbb{R}^{n_{\boldsymbol{u}}\times n_{\boldsymbol{v}}}, \mathbb{R}^{n_{\boldsymbol{u}}})$, $g \in \mathcal{C}^\infty(\mathbb{R}^{n_{\boldsymbol{u}}\times n_{\boldsymbol{v}}}, \mathbb{R}^{n_{\boldsymbol{v}}})$, with $n_{\boldsymbol{v}}, n_{\boldsymbol{u}}$ being positive integers. Also, assume that there is a probability measure $\mu(\mathrm{d}\boldsymbol{u},\mathrm{d}\boldsymbol{v}) = \rho(\boldsymbol{u},\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}$ with smooth density $\rho \in \mathcal{C}^\infty(\mathbb{R}^{n_{\boldsymbol{u}}\times n_{\boldsymbol{v}}}, [0,\infty))$, which can be associated with a stationary state [2] of the system (1.31). Consider now the projection operator $\mathscr{P}$, which maps observables $w(\,\cdot\,,\,\cdot\,)$ onto the conditional expectation $\mathscr{P}\boldsymbol{u} \mapsto \mathbb{E}_\mu[w(\boldsymbol{u},\boldsymbol{v}) \mid \boldsymbol{v}]$, i.e.,

$$(\mathscr{P}w)(\boldsymbol{u}) = \frac{\int_{\mathbb{R}^{n_{\boldsymbol{v}}}} \rho(\boldsymbol{u},\boldsymbol{v})w(\boldsymbol{u},\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}}{\int_{\mathbb{R}^{n_{\boldsymbol{v}}}} \rho(\boldsymbol{u},\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}}.$$

The Mori-Zwanzig projection formalism allows to formally rewrite the system (1.31) as an integro-differential equation of the generic form

$$\dot{\boldsymbol{u}}(t) = \bar{f}(\boldsymbol{u}(t)) + \int_0^t K(\boldsymbol{u}(t-s),s)\mathrm{d}s + \eta(\boldsymbol{u}(0),\boldsymbol{v}(0),t), \tag{1.33}$$

where $\bar{f} = \mathscr{P}f$, $K : \mathbb{R}^{n_{\boldsymbol{u}}} \times [0,\infty) \to \mathbb{R}^{n_{\boldsymbol{u}}}$ is a memory kernel, and η is a function of the initial values of $\boldsymbol{u},\boldsymbol{v}$ and the time variable t. It is important to note that while η depends on the initial condition of both $\boldsymbol{u}$ and $\boldsymbol{v}$ in (1.31), the remaining terms in the integro-differential equation (1.33) only depend explicitly on the dynamic variable $\boldsymbol{u}$. Similarly as in the stochastic integro-differential equations (1.28) and (1.30) the convolution term in (1.33) can be under appropriate conditions of f, g considered as a dissipation term. Likewise, under the assumptions that $\boldsymbol{u},\boldsymbol{v}$ are initialised randomly according to μ, the term $\eta(\boldsymbol{u}(0),\boldsymbol{v}(0),t)$ in (1.33) can be interpreted as a random force

[2] in the sense that $\mathcal{L}\rho = 0$, with $\mathcal{L}$ being the Liouville operator associated with (1.31)

/ fluctuating term.

Of particular interest in our case is the situation where the functions f, g are such that $(f^T, g^T)^T$ is a Hamiltonian vector field and (1.31) corresponds to the equation of motion of a Hamiltonian system. In this case a natural choice for μ is the Gibbs-Boltzmann distribution associated with the Hamiltonian. This choice of μ allows to interpret the degrees of freedom represented by the dynamical variable $\boldsymbol{v}$ as a heat bath or energy reservoir. For example, let $\boldsymbol{u} = (\boldsymbol{q}, \boldsymbol{p}) \in \mathbb{R}^{2n}$, $\boldsymbol{v} = (\tilde{\boldsymbol{q}}, \tilde{\boldsymbol{p}}) \in \mathbb{R}^{2m}$ with $2n = n_{\boldsymbol{u}}, 2m = n_{\boldsymbol{v}}$. We may consider the case where f, g are derived from the Hamiltonian

$$H(\boldsymbol{q}, \boldsymbol{p}, \tilde{\boldsymbol{q}}, \tilde{\boldsymbol{p}}) = V(\boldsymbol{q}) + \frac{1}{2}\boldsymbol{p}^T \boldsymbol{M} \boldsymbol{p} + V_c(\boldsymbol{q}, \tilde{\boldsymbol{q}}) + V_h(\tilde{\boldsymbol{q}}) + \frac{1}{2}\tilde{\boldsymbol{p}}^T \widetilde{\boldsymbol{M}} \tilde{\boldsymbol{p}}, \tag{1.34}$$

where V, V_c, V_h are smooth potential functions so that $V + V_c + V_h$ is confining and $\boldsymbol{M} \in \mathbb{R}^{n \times n}, \widetilde{\boldsymbol{M}} \in \mathbb{R}^{m \times m}$ are symmetric positive definite matrices. In the view of (1.33) the variables $(\boldsymbol{q}, \boldsymbol{p})$ correspond to the explicitly resolved part of the system; the variables $(\tilde{\boldsymbol{q}}, \tilde{\boldsymbol{p}})$ correspond to the part of the system which is "projected out" and is replaced with the dissipation term and the fluctuation term, thus functions as the heat bath in the reduced model. The coupling between heat bath and explicitly resolved degrees of freedom is encoded in the form of the coupling potential V_c, and the statistical properties of the heat bath are both determined by the form of the mass matrix $\widetilde{\boldsymbol{M}}$ and the form of the potential V_h.

Let P denote the projection $(\boldsymbol{u}, \boldsymbol{v}) \mapsto \boldsymbol{u}$. The first step in the derivation of the integro-differential equation (1.33) is to rewrite the first line in (1.31) as

$$\dot{\boldsymbol{u}}(t) = (\mathscr{P} f)\left(P(\boldsymbol{u}(t), \boldsymbol{v}(t))\right) + \left[f(\boldsymbol{u}(t), \boldsymbol{v}(t)) - (\mathscr{P} f)\left(P(\boldsymbol{u}(t), \boldsymbol{v}(t))\right)\right]. \tag{1.35}$$

Obviously, the first term in (1.35) corresponds exactly to $\bar{f}(\boldsymbol{u}(t))$ in (1.33). Let

$$\mathcal{L} = f(\boldsymbol{u}, \boldsymbol{v}) \cdot \nabla_{\boldsymbol{u}} + g(\boldsymbol{u}, \boldsymbol{v}) \cdot \nabla_{\boldsymbol{v}}$$

denote the Liouville operator associated with (1.31). Noting that

$$\mathcal{L}\left(P(\boldsymbol{u}, \boldsymbol{v})\right) = f(\boldsymbol{u}, \boldsymbol{v})$$

the term in the square brackets in (1.33) can be rewritten in semi-group notation as

$$\begin{aligned} f(\boldsymbol{u}(t), \boldsymbol{v}(t)) - (\mathscr{P} f)\left(\boldsymbol{u}(t), \boldsymbol{v}(t)\right) &= e^{t\mathcal{L}}(\boldsymbol{I} - \mathscr{P}) f(\boldsymbol{u}(0), \boldsymbol{v}(0)) \\ &= e^{t\mathcal{L}}(\boldsymbol{I} - \mathscr{P}) \mathcal{L} P(\boldsymbol{u}(0), \boldsymbol{v}(0)), \end{aligned} \tag{1.36}$$

where $e^{t\mathcal{L}}$ denotes the flow-map operator associated with the solution of (1.31), which is defined so that $e^{t\mathcal{L}} w(\boldsymbol{u}(0), \boldsymbol{v}(0)) = w(\boldsymbol{u}(t), \boldsymbol{v}(t))$. The integro-differential form (1.33) then follows by applying the operator identiy

$$e^{t\mathcal{L}} = \int_0^t e^{(t-s)\mathcal{L}} \mathscr{P} \mathcal{L} e^{s(\boldsymbol{I} - \mathscr{P})\mathcal{L}} \mathrm{d}s + e^{t(\boldsymbol{I} - \mathscr{P})\mathcal{L}},$$

which is known as Dyson's formula [89], to the last line in (1.36) yielding

$$e^{t\mathcal{L}}(\boldsymbol{I}-\mathscr{P})\mathcal{L}P(\boldsymbol{u}(0),\boldsymbol{v}(0)) = \int_0^t e^{(t-s)\mathcal{L}}\mathscr{P}\mathcal{L}e^{s(\boldsymbol{I}-\mathscr{P})\mathcal{L}}(\boldsymbol{I}-\mathscr{P})\mathcal{L}P(\boldsymbol{u}(0),\boldsymbol{v}(0))\mathrm{d}s + e^{t(\boldsymbol{I}-\mathscr{P})\mathcal{L}}(\boldsymbol{I}-\mathscr{P})\mathcal{L}P(\boldsymbol{u}(0),\boldsymbol{v}(0)), \tag{1.37}$$

where the second term on the right hand side can be identified with η in (1.33), and the first term in (1.37) corresponds to the integral term in (1.33). The form of the last term in (1.37) suggests, that η can be formally written as the solution of a differential equation of the form

$$\begin{aligned}\frac{\partial}{\partial t}\eta(\boldsymbol{u}(0),\boldsymbol{v}(0),t) &= (\boldsymbol{I}-\mathscr{P})\mathcal{L}\eta(\boldsymbol{u}(0),\boldsymbol{v}(0),t),\\ \eta(\boldsymbol{u}(0),\boldsymbol{v}(0),0) &= f(\boldsymbol{u}(0),\boldsymbol{v}(0)) - (\mathscr{P}f)(\boldsymbol{u}(0)),\end{aligned} \tag{1.38}$$

which is commonly referred to as the *orthogonal dynamics equation* [26, 41].

A couple of remarks are in order. First, we reiterate that the above calculations are purely formal, i.e., the above expressions for the memory kernel K and the fluctuation term η do in general not posses a closed form solution and are therefore often considered as intractable in their exact form. Moreover, the well-posedness of the orthogonal dynamics equation (1.38) is not obvious and care needs to be taken regarding the existence of solutions and the interpretation of the differential operator $\mathcal{L}$ therein. We refer in this regard to [40] for a rigorous study of this equation. We also mention that the above choice of the projection operator $\mathscr{P}$ as a linear operator which maps functions of $(\boldsymbol{u},\boldsymbol{v})$ into the space of functions of $\boldsymbol{u}$ constitutes as such a special case within the Mori-Zwanzig formalism. More general forms of the projection operator $\mathscr{P}$ can be considered within the Mori-Zwanzig formalism. For example, the Mori-Zwanzig formalism can be used to formally derive an integro-differential equation for the dynamics of reaction coordinates / collective variables. The corresponding projection operator $\mathscr{P}$ is typically non-linear in these cases, which can drastically complicate the derivation and the form of the integro-differential equation. For a more general presentation of the Mori-Zwanzig projection formalism we refer to the above mentioned papers [26, 41] and the references therein as well as the original papers by Mori [88] and Zwanzig [130, 131]. In particular the latter paper by Zwanzig considers non-linear forms of the projection operator $\mathscr{P}$.

Secondly, we point out that in order to derive the stochastic integro-differential equations (1.28) and (1.30) yet another step is required. While (1.28) and (1.30) are of the form of a stochastic integro-differential equation, i.e., they are integro-differential equations driven by a (non-Markovian) stochastic process, the equation (1.33) constitutes an integro-differential equation with random initial data, i.e., the system follows a deterministic trajectory after initialisation. In the physics literature it is common in the situation where f, g form a Hamiltonian vector field to establish equivalence of these systems by virtue of an averaging argument which is considered valid when the system is in equilibrium and $n_{\boldsymbol{v}}$ is sufficiently large (see e.g. [60]). Drawing a mathematically rigorous connection between (1.33) and a suitable stochastic integro-differential equation which resembles the form of (1.28) or (1.30) requires substantial work. As we discuss in the section below, weak convergence as $n_{\boldsymbol{v}} \to \infty$ of the trajectory of $\boldsymbol{u}$ on finite time intervals to the solution of a stochastic integro-differential has been shown

in [66, 65] for instances of the Ford-Kac model.

1.3.2 Derivation of the generalised Langevin equation from an infinite dimensional heat-bath

The rather cumbersome step of deriving a stochastic integro-differential equation from the integro-differential equation (1.33) in the limit $n_{\boldsymbol{v}} \to \infty$ can be elegantly circumvented by coupling the variable $\boldsymbol{u} = (\boldsymbol{q}, \boldsymbol{p})$ to a heat bath which is of the form of an infinite dimensional Hamiltonian system. In what follows we outline this alternative approach of deriving the stochastic integro-differential equation (1.28) closely following the presentations in [101] and [98, Chapter 8]. We emphasise that the presentation is far from being self-contained and we refer for a detailed presentation to the above mentioned references.

A natural choice for an infinite dimensional heat-bath model is the linear wave equation

$$\begin{aligned} \frac{\partial}{\partial t}\phi &= \pi, \\ \frac{\partial}{\partial t}\pi &= \left(\frac{\partial}{\partial x}\right)^2 \phi, \end{aligned} \tag{1.39}$$

which can be considered as an infinite-dimensional Hamiltonian system with associated Hamiltonian

$$H_b(\phi, \pi) = \int_{\mathbb{R}} |\partial_x \phi(x)|^2 + |\pi(x)|^2 \mathrm{d}x, \tag{1.40}$$

It can be shown that solutions of the wave equation (1.39) can be considered as elements of a Hilbert space $\mathscr{H}$, whose inner product induces a norm which corresponds to the Hamiltonian $H_b(\phi, \pi)$. The coupling of the wave equation with the variable $\boldsymbol{u} = (\boldsymbol{q}, \boldsymbol{p})$ is modelled by a density function ρ so that the integral term

$$\boldsymbol{q} \int_{\mathbb{R}} \rho(x) \partial_x \phi(x) \mathrm{d}x, \tag{1.41}$$

can be interpreted as a coupling potential. Together with the Hamiltonian

$$H(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^2 + V(\boldsymbol{q}), \ (\boldsymbol{q}, \boldsymbol{p}) \in \mathbb{R}^2,$$

where $V \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$ is assumed to be a confining potential, the combined Hamiltonian of the whole system then takes the form

$$\mathcal{H}(\boldsymbol{q}, \boldsymbol{p}, \phi, \pi) = \frac{1}{2}\boldsymbol{p}^2 + V(\boldsymbol{q}) + H_b(\phi, \pi) + \boldsymbol{q} \int_{\mathbb{R}} \rho(x) \partial_x \phi(x) \mathrm{d}x. \tag{1.42}$$

From the Hamiltonian (1.42) the equation of motions of the combined system can be derived. Under the assumption that that the initial states of ϕ, π are distributed according to the Gibbs measure[3] associated with the Hamiltonian $\mathcal{H}$, it can then be shown that the solution of the equation of motion can be written in the form of the

[3]Note that such a measure can indeed be explicitly constructed on $\mathbb{R}^2 \times \mathscr{H}$. This follows since the Hamiltonian $\mathcal{H}$ is a quadratic functional in ϕ, π, which means that the Gibbs measure conditioned on the values of $\boldsymbol{q}, \boldsymbol{p}$ can be considered as a Gaussian measure on the Hilbert space $\mathscr{H}$. A comprehensive review of Gaussian measures on Hilbert spaces can for example be found in [110, Appendix C].

stochastic integro-differenital equation (1.28) with

$$U(\boldsymbol{q}) = V(\boldsymbol{q}) + \frac{\lambda}{2}\boldsymbol{q}^2,$$

and

$$\boldsymbol{K}(t) = \int_{\mathbb{R}} |\hat{\rho}(k)|^2 e^{ikt} \mathrm{d}k,$$

where $\hat{\rho}(k)$ denotes the Fourier transform of ρ and $\lambda = \int_{\mathbb{R}} |\rho(x)|^2 \mathrm{d}x$. Generalised Langevin equations derived from a Hamiltonian of the form (or similar to) (1.42) have been extensively studied in [53, 54, 55]. Likewise, the (non-equilibrium) models by Rey-Bellet and coworkers (see e.g. [35, 34, 100, 101]) are derived from a Hamiltonian similar to (1.42).

1.3.3 The Ford-Kac model

We consider again the Mori-Zwanzig projection formalism in the situation where the ODE (1.31) corresponds to the equation of motion derived from the Hamiltonian (1.34). We already mention above that the memory kernel K and the fluctuation term in the integro-differential equation (1.33) do in general not possess a closed form solution. A notable exception, however, is the situation of a linearly coupled harmonic heat bath, e.g.,

$$V_c(\boldsymbol{q}, \tilde{\boldsymbol{q}}) = \boldsymbol{q}^T \Omega_c \tilde{\boldsymbol{q}}, \tag{1.43}$$

with $\Omega_c \in \mathbb{R}^{n\times m}$, and

$$V_h(\tilde{\boldsymbol{q}}) = \frac{1}{2}\tilde{\boldsymbol{q}}^T \Omega_h \tilde{\boldsymbol{q}}, \tag{1.44}$$

with $\Omega_h \in \mathbb{R}^{m\times m}$ being a symmetric positive (semi-)definite matrix. Under this choice of the potential functions V_c and V_h, the equation of motion associated with (1.34) is of the form

$$\begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{M}^{-1}\boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= -\nabla_{\boldsymbol{q}} V(\boldsymbol{q}) + \Omega_c \tilde{\boldsymbol{q}}, \\ \dot{\tilde{\boldsymbol{q}}} &= \widetilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{p}}, \\ \dot{\tilde{\boldsymbol{p}}} &= -\Omega_h \tilde{\boldsymbol{q}} + \Omega_c^T \boldsymbol{q}. \end{aligned} \tag{1.45}$$

The system (1.45) was first studied in [37] and is commonly referred to as *Ford-Kac model*. Integrating the 3rd and 4th line of (1.45) we obtain

$$\begin{pmatrix} \tilde{\boldsymbol{q}}(t) \\ \tilde{\boldsymbol{p}}(t) \end{pmatrix} = e^{t\boldsymbol{R}} \begin{pmatrix} \tilde{\boldsymbol{q}}(0) \\ \tilde{\boldsymbol{p}}(0) \end{pmatrix} + \int_0^t e^{(t-s)\boldsymbol{R}} \begin{pmatrix} \boldsymbol{0} \\ \Omega_c^T \boldsymbol{q}(s) \end{pmatrix} \mathrm{d}s, \tag{1.46}$$

where by $\boldsymbol{R} \in \mathbb{R}^{2m\times 2m}$ we denote the matrix

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{0} & \widetilde{\boldsymbol{M}}^{-1} \\ -\Omega_h & 0 \end{pmatrix}.$$

Partial integration of the integral term in (1.46) yields

$$\begin{pmatrix} \tilde{\boldsymbol{q}}(t) \\ \tilde{\boldsymbol{p}}(t) \end{pmatrix} = e^{t\boldsymbol{R}} \begin{pmatrix} \tilde{\boldsymbol{q}}(0) \\ \tilde{\boldsymbol{p}}(0) \end{pmatrix} + \boldsymbol{R}^{-1} \begin{pmatrix} \boldsymbol{0} \\ \Omega_c^T \boldsymbol{q}(t) \end{pmatrix} - \boldsymbol{R}^{-1} e^{t\boldsymbol{R}} \begin{pmatrix} \boldsymbol{0} \\ \Omega_c^T \boldsymbol{q}(0) \end{pmatrix} + \int_0^t e^{(t-s)\boldsymbol{R}} \begin{pmatrix} \boldsymbol{0} \\ \Omega_c^T \boldsymbol{p}(s) \end{pmatrix} \mathrm{d}s.$$

Substituting $\tilde{\boldsymbol{q}}$ in the 2nd line by this expression we obtain an integro-differential equation of the form (1.33) with the deterministic vector field $\bar{f}$ being of the form

$$\bar{f}(\boldsymbol{q},\boldsymbol{p}) = \begin{pmatrix} \boldsymbol{M}^{-1}\boldsymbol{p} \\ -\nabla_{\boldsymbol{q}} V(\boldsymbol{q}) - \Omega_c \Omega_h \Omega_c^T \boldsymbol{q} \end{pmatrix},$$

the memory kernel K being of the form

$$K(\boldsymbol{p}(t-s),s) = -\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \Omega_c^{-1} \end{pmatrix} e^{(t-s)\boldsymbol{R}} \begin{pmatrix} \boldsymbol{0} \\ \Omega_c^T \boldsymbol{p}(s) \end{pmatrix}, \tag{1.47}$$

and the fluctuation term being of the form

$$\eta(\tilde{\boldsymbol{q}}(0),\tilde{\boldsymbol{p}}(0),\boldsymbol{q}(0),t) = e^{t\boldsymbol{R}} \begin{pmatrix} \tilde{\boldsymbol{q}}(0) \\ \tilde{\boldsymbol{p}}(0) \end{pmatrix} - \boldsymbol{R}^{-1} e^{t\boldsymbol{R}} \begin{pmatrix} \boldsymbol{0} \\ \Omega_c^T \boldsymbol{q}(0) \end{pmatrix}. \tag{1.48}$$

1.3.4 The thermodynamic limit of the Ford-Kac model

A detailed analysis of the thermodynamic limit $m \to \infty$ of an instance of the Ford-Kac model can be found in [66]; see also [65, 41]. The Hamiltonian of the system considered in [66] comprises a single distinguished particle of unit mass, which is subject to an external force associated with the confining potential function $U \in \mathcal{C}^\infty(\mathbb{R},\mathbb{R})$. The heat bath is modelled by m particles. Each of the heat bath particles is attached by a linear spring to the distinguished particle. The heat bath particles are not subject to any additional force apart from the coupling force. The corresponding Hamiltonian can be written as

$$H(\boldsymbol{q},\boldsymbol{p},\tilde{\boldsymbol{q}},\tilde{\boldsymbol{p}}) = \frac{1}{2}\boldsymbol{p}^2 + U(\boldsymbol{q}) + \frac{1}{2}\sum_{j=1}^m \frac{\tilde{\boldsymbol{p}}_j^2}{\tilde{m}_j} + \frac{1}{2}\sum_{j=1}^m k_j(\tilde{\boldsymbol{q}}_j - \boldsymbol{q}),^4 \tag{1.49}$$

where $k_j > 0$ corresponds to the stiffness constant of the spring attached to the j-th heat bath particle and $\tilde{m}_j > 0$ is the mass of the j-th heat bath particle. For this system one finds that the terms (1.47) and (1.48) take a particular simple form, so that the corresponding integro-differential equation can be written as

$$\begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= -\partial_{\boldsymbol{q}} U(\boldsymbol{q}) - \int_0^t K^{(m)}(t-s)\boldsymbol{p}(s)\mathrm{d}s + \eta^{(m)}(\tilde{\boldsymbol{q}}_i,\tilde{\boldsymbol{p}}_i,t), \end{aligned} \tag{1.50}$$

where the memory kernel is of the form

$$K^{(m)}(t) = \sum_{i=1}^m k_i \cos(\omega_i t),$$

and the fluctuation term is of the form

$$\eta^{(m)}(\tilde{\boldsymbol{q}}_i,\tilde{\boldsymbol{p}}_i,t) = \sum_{i=1}^m \sqrt{\frac{k_i}{\beta}} \Big(\tilde{\boldsymbol{q}}_i(0)\cos(\omega_i t) + \tilde{\boldsymbol{p}}_i(0)\sin(\omega_i t) \Big),$$

[4] One easily verifies that this Hamiltonian corresponds to a parametrisation of (1.34) as $\boldsymbol{M} = 1$, $\widetilde{\boldsymbol{M}} = \mathrm{diag}(\tilde{m}_1,\ldots,\tilde{m}_m)$, $V(\boldsymbol{q}) = U(\boldsymbol{q}) + \frac{1}{2}\sum_{i=1}^m k_i \boldsymbol{q}^2$, $V_c(\boldsymbol{q},\tilde{\boldsymbol{q}}) = \sum_{i=1}^m k_i \boldsymbol{q}\tilde{\boldsymbol{q}}_i$, $V_h(\tilde{\boldsymbol{q}}) = \frac{1}{2}\sum_{i=1}^m k_i \tilde{\boldsymbol{q}}_i^2$.

with $\omega_j = \sqrt{k_j/\tilde{m}_j}$. If the initial conditions of the heat bath particles are assumed to be distributed according to the Gibbs-measure associated with (1.49) and the statistical distribution of the values of k_j and $\tilde{m}_j$ are controlled in a certain way as $m \to \infty$, it can been shown that for any finite $T > 0$ the trajectories of the solution of (1.50) converges weakly within the interval $[0, T]$ to the solution of a stochastic integro-differential equation of the form (1.28); for a precise statement see [66, Theorem 4.1].

1.3.5 The Kac-Zwanzig model

The Kac-Zwanzig model (see [131]) is a generalisation of the Ford-Kac model, the heat bath is still harmonic, i.e., V_h is of the from (1.44), but the coupling potential is such that the coupling force is linear in $\tilde{\boldsymbol{q}}$ but non-linear in $\boldsymbol{q}$, i.e.,

$$V_c(\boldsymbol{q}, \tilde{\boldsymbol{q}}) = \boldsymbol{G}(\boldsymbol{q})\tilde{\boldsymbol{q}},$$

where $\boldsymbol{G} \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^{n\times m})$. For such a system a closed form solution of the terms in the Mori-Zwanzig projection (1.33) can still be derived (see [131] or [46] for a detailed derivation). However, unlike in the situation of the Ford-Kac model the closed form solution of the memory kernel K and the fluctuation term η are functions of $\boldsymbol{q}$. This observation motivates to consider GLEs of the form (1.30). Instances of (1.30) which are derived from such a Kac-Zwanzig heat bath model can be found for example in [60, 115, 93, 94].

1.4 Original contributions of this Thesis

I detail here the novel results presented in this thesis:

(i) new conditions for ergodicity of GLEs: in previous work (see [96]) exponential convergence in law was shown for GLEs with exponential memory kernels and no cross-correlation. We extend these results to cover GLEs with memory kernels of significantly more general forms.

(ii) numerical methods: we systematically investigate different splitting strategies for the Markovian representation of quasi-Markovian GLEs and show that the corresponding discretised processes are geometrically ergodic under appropriate conditions. We also propose modifications of the above mentioned schemes which allow an efficient integration of the fluctuation-dissipation part in GLEs with non-stationary noise or memory kernels with large numbers of cross-correlation terms.

(iii) applications in sampling: We demonstrate that one of the proposed GLE schemes leads to a drastic reduction of the error due to discretisation in ergodic averages in comparison to other comparable GLE schemes previously proposed in the literature. We show analytically that for this scheme the discretisation bias of ergodic averages of configurational observables vanished in the case of a harmonic force. We show formally that this scheme posses a super-convergence property in the over-damped limit. In numerical experiments we find that this scheme when combined with a pre-optimised memory kernel as proposed in [22] performs favourably in comparison to comparable schemes based on an underdamped Langevin diffusion process.

(iv) adaptive GLE schemes: We propose a novel adaptive thermostat method based on the GLE which automatically corrects for time-correlated gradient noise and thereby ensures accurate sampling of a prescribed target density. We show that this method can be applied in the context of large scale Bayesian inference.

All the above mentioned results are joint work with my supervisor Prof. Benedict Leimkuhler. (iv) is joint work with Prof. Jianfeng Lu, Duke University; and Mark Rowland, University of Cambridge. Chapter 5 and parts of Section 1.2.4 are based on a manuscript written in the course of that collaboration. Chapter 2 in the revised version of this thesis contains passages which were also used in the article [108], which was written in collaboration with Prof. Benedict Leimkuhler and Prof. Vincent Danos, École Normale Supérieure, Paris. The possibility of showing the minorisation condition via Girsanov's theorem in the proof of Lemma 3.4.7 was pointed out to me by Jonathan Mattingly, Duke University.

The remainder of this thesis is structured as follows. In the next chapter we revise some basic results from stochastic analysis, ergodic theory for Markov process, and the construction of stochastic splitting schemes, which we will need for the derivation of our results in later sections. In Chapter 3 we introduce the extended variable formalism for quasi-Markovian GLEs which we use throughout this thesis. In Section 3.4 we derive results on the ergodicity of quasi-Markovian GLEs and revise results on overdamped and the white-noise limit of quasi-Markovian GLEs. In Chapter 4 we discuss numerical integration strategies for the extended variable formalism. In Chapter 5 we introduce an adaptive thermostat which is based on the GLE and which allows accurate sampling in the presence of (time-)correlated gradient noise.

Chapter 2

Sampling via stochastic differential equations

In this chapter we revise some basic concepts from stochastic analysis, ergodic theory for Markov processes, and the construction of stochastic splitting schemes.

2.1 Notation and basic concepts

Let $X = (X(t))_{t\in\mathcal{T}}$ be a Markov process with associated probability space $(\Omega, \mathscr{A}, \mathbb{P})$ and statespace $(\mathscr{B}(\mathbb{R}^n), \mathbb{R}^n)$, where $\mathscr{B}(\mathbb{R}^n)$ denotes the Borel σ-algebra induced by the Euclidian norm $\|\cdot\|_2$ on $\mathbb{R}^n$. We consider both the case of a continuous time Markov processes, where $\mathcal{T} = [0, \infty)$ and the case of a discrete time Markov process, where $\mathcal{T} = \mathbb{N}$. In the latter, discrete time case, we will occasionally use index notation $X_k = X(k)$. We denote by $(\mathcal{P}_t)_{t\in\mathcal{T}}$ the semigroup of operators which describe the evolution of the expectation of test functions $\varphi \in \mathscr{S} \subset \{\varphi : \mathbb{R}^n \to \mathbb{R}, \text{measurable}\}$, i.e.,

$$\mathcal{P}_t\varphi(x) := \mathbb{E}^x[\varphi(X(t))].$$

where $\mathbb{E}^x\varphi(X(t))$ is used here and in the sequel as a short hand notation for the conditional expectation $\mathbb{E}[\varphi(X(t))|X(0) = x]$. Similarly, by $(\mathcal{P}_t^\dagger)_{t\in\mathcal{T}}$, we denote the semigroup describing the evolution of the law $\mathbb{P}_{X(t)}, t \in \mathcal{T}$ of the process X, i.e.,

$$\mathcal{P}_t^\dagger \mathbb{P}_{X(0)} := \mathbb{P}_{X(t)}.$$

The semigroup $(\mathcal{P}_t^\dagger)_{t\in\mathcal{T}}$ and $(\mathcal{P}_t)_{t\in\mathcal{T}}$ are formally adjoint, i.e.,

$$\int_{\Omega_{\boldsymbol{x}}} \mathcal{P}_t\varphi(x)\mathbb{P}_{X(0)}(\mathrm{d}x) = \int_{\Omega_{\boldsymbol{x}}} \varphi(x)\mathcal{P}_t^\dagger\mathbb{P}_{X(0)}(\mathrm{d}x).$$

Furthermore, it follows directly from the definition of the semigroup $(\mathcal{P}_k^\dagger)_{k\in\mathbb{N}}$ that the transition kernel $P(x, \mathrm{d}y)$ associated with the Markov chain $(X_k)_{k\in\mathbb{N}}$ takes the form

$$P(x, \mathrm{d}y) = \mathcal{P}_1^\dagger\delta_x(\mathrm{d}y).$$

Ergodic Markov processes

A probability measure $\pi(\mathrm{d}x) : \mathscr{B}(\mathbb{R}^n) \to [0,1]$ is said to preserved by Markov process X if it is invariant under the action of the semigroup $\mathcal{P}_t^\dagger$, i.e.,

$$\mathcal{P}_t^\dagger \pi(\mathrm{d}x) = \pi(\mathrm{d}x), \text{ for all } t \in \mathcal{T}.$$

where $\mathcal{T} = \mathbb{N}$ for the discrete time case and $\mathcal{T} = [0,\infty)$ for the continuous time case. In this case we also say that π is an *invariant measure* of the process X. For a given Markov process, the set of invariant measures is convex, i.e. any convex combination of two invariant measures is an invariant measure. Let $\mathcal{M}(\mathbb{R}^n, \mathbb{R})$ denote the set of Borel-measurable real-valued functions on $\mathbb{R}^n$. By

$$\mathbb{E}_\pi : \mathcal{M}(\mathbb{R}^n, \mathbb{R}) \to \mathbb{R} \cup \{-\infty, \infty\},$$

we denote the operator, which maps measurable functions onto their π-weighted average, i.e.,

$$\mathbb{E}_\pi : \varphi \mapsto \int_{\Omega_{\boldsymbol{x}}} \varphi(x)\pi(\mathrm{d}x).$$

Let

$$L^1(\pi) := \{\varphi \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}) : \mathbb{E}_\pi|\varphi| < \infty\}.$$

The process X is said to be *ergodic* with respect to $\pi(\mathrm{d}x)$, if for all $\varphi \in L^1(\pi)$, trajectory averages converge almost surely to expectations with respect to the probability measure π, i.e.,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \varphi(X(\omega,t))\mathrm{d}t = \mathbb{E}_\pi \varphi, \tag{2.1}$$

for all $\varphi \in L^1(\pi)$ and for almost all $\omega \in \Omega$, if $\mathcal{T} = [0,\infty)$, and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \varphi(X(\omega,n)) = \mathbb{E}_\pi \varphi, \tag{2.2}$$

for all $\varphi \in L^1(\pi)$ and for almost all $\omega \in \Omega$, if $\mathcal{T} = \mathbb{N}$. Conversely, if the above properties hold for a Markov process X, then we call π an *ergodic measure*. If π is an ergodic measure of X, then one easily verifies that π is an invariant measure of X and as such unique.

2.2 Stochastic integration

Let $\big(W(t)\big)_{t\in[0,\infty)}$ be a Wiener processes defined on the probability space $(\Omega, \mathscr{A}, \mathbb{P})$ i.e.,

(i) $W(0) = 0$,

(ii) sample paths $t \mapsto W(\omega,t)$, are continuous for $\mathbb{P}$-almost all $\omega \in \Omega$,

(iii) W has independent and increments,

(iv) $W(t) - W(s) \sim \mathcal{N}(0, t-s)$ for $0 \le s \le t$.

Let $\big(f(t)\big)_{t\in[0,\infty)}$ be a stochastic process adapted to the natural filtration of the

Wiener process W and assume that f is square integrable in the sense

$$\mathbb{E}\left(\int_0^t f(s)^2 \mathrm{d}s\right) < \infty.$$

For fixed $T > 0$ and $N \in \mathbb{N}$ denote by $t_k = \frac{Tk}{K-1}, k = 1, \ldots, K-1$ the equidistant discretisation of the time interval $[0, T]$. For $\lambda \in [0, 1]$, define the point set $\tau_k = (1-\lambda)t_k + \lambda t_{k+1}, k = 1, \ldots, K-1$. For this point set we denote by

$$\mathcal{I}_{\lambda,T,K}[f, W] = \sum_{k=0}^{K-1} f(\tau_k)(W(\tau_{k+1}) - W(\tau_k)),$$

the corresponding W-weighted Riemann sum, and define accordingly stochastic integrals as the limit

$$\mathcal{I}_{\lambda,T}[f, W] = \lim_{K\to\infty} \mathcal{I}_{\lambda,T,K}[f, W]. \tag{2.3}$$

For details regarding the existence of the limit (2.3) and the functional space in which the convergence of $\mathcal{I}_{\lambda,T,K}[f, W]$ is to be understood we refer to e.g. [95, Chapter 3]. It is important to note that the value of $\mathcal{I}_{\lambda,T}[f, W]$ does in general depend on the the choice of λ. The most common choices for λ are either $\lambda = 0$ or $\lambda = 1/2$. In the former case $\mathcal{I}_{\lambda,T}[f, W]$ corresponds to the definition of the *Itô integral*. For $\lambda = 1/2$ the limit (2.3) corresponds to the definition of the *Stratonovich integral*.

2.3 Stochastic differential equations

Let $\boldsymbol{a} : \mathbb{R}^n \to \mathbb{R}^n$ and $\boldsymbol{b} : \mathbb{R}^n \to \mathbb{R}^{n\times n}$, be measurable functions and $W = (W_1, \ldots, W_n)$ a vector of n independent Wiener processes. A stochastic differential equation (SDE) is typically denoted as

$$\mathrm{d}X = \boldsymbol{a}(X)\mathrm{d}t + \boldsymbol{b}(X)\mathrm{d}W,\ X(0) \sim \mu_0, \tag{2.4}$$

which is the shorthand form of the integral equation,

$$X(t) - X(0) = \int_0^t \boldsymbol{a}(X(s))\mathrm{d}s + \int_0^t \boldsymbol{b}(X(s))\mathrm{d}W(s). \tag{2.5}$$

The probability measure μ_0 defines the initial distribution of the process $X(0)$. Throughout the remainder of this chapter we will assume

$$\int_{\mathbb{R}^n} \|x\|_2^2 \mu_0(\mathrm{d}x) < \infty.$$

For sufficiently regular $\boldsymbol{a}$, the first integral in (2.5), is to be interpreted as a Riemann integral, whereas the stochastic integral,

$$\int_0^t \boldsymbol{b}(X(s))\mathrm{d}W(s),$$

may be either interpreted as a Itô integral or a Stratonovich integral, i.e.,

$$\int_0^t \boldsymbol{b}(X(s))\mathrm{d}W(s) = \mathcal{I}_{\lambda,t}[\boldsymbol{b}(X), W]$$

with $\lambda = 0$ or $\lambda = 1/2$, respectively.

If not stated otherwise we interpret throughout this thesis stochastic integrals as being defined in the Itô sense and in cases where a Stratonovich interpretation of the stochastic integral is appropriate we indicate that by writing

$$\boldsymbol{b}(X) \circ \mathrm{d}W(t),$$

as the shorthand notation for the stochastic part in (2.4). Interpreting the stochastic term in the Itô sense, allows us to assume the validity of the following identities:

Itô isometry. Let f be defined as above. Then

$$\mathbb{E}\left[\left(\int_0^t f(s)\mathrm{d}W(s)\right)^2\right] = \mathbb{E}\left[\int_0^t \left(f(s)\right)^2 \mathrm{d}s\right], \tag{2.6}$$

where $\mathbb{E}$ denotes the expectation with respect to white noise process W or more specifically with respect to the classical Wiener measure (See e.g. [11]).

Itô Doeblin-formula (Itô's Lemma). Let now $f : \mathbb{R}^n \to \mathbb{R}$ denote a twice differentiable function, and X to be an Itô diffusion process satisfying the SDE (2.4), then the process $f(X)$ satisfies the SDE

$$\mathrm{d}f(X) = \left(\boldsymbol{a}(X) \cdot \nabla f(X) + \frac{1}{2}\boldsymbol{b}(X)\boldsymbol{b}^\top(X) : \nabla^2 f(X)\right)\mathrm{d}t + \nabla f^\top(X)\boldsymbol{b}(X)\mathrm{d}W(t),$$

where $\nabla^2 f(X)$ denotes the Hessian of the function f evaluated at X, and ' $\cdot$ ' denotes the inner product operator, i.e.,

$$\boldsymbol{a} \cdot \nabla f = \sum_{1 \le i \le n} \boldsymbol{a}_i f_i,$$

and ' : ' denotes the Frobenius product operator, i.e.,

$$\boldsymbol{b}^\top \boldsymbol{b} : \nabla^2 f = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \boldsymbol{b}_{i,k}\boldsymbol{b}_{j,k}\partial_{x_i,x_j} f.$$

Furthermore, the definition of the Itô integral implies a very useful martingale property, which is summarised in the following proposition.

Proposition 2.3.1. *Let $f(X)$ be stochastic process adapted to the filtration of the Wiener process W. Let $T > 0$. If f is square integrable in the sense that*

$$\mathbb{E}\left[\int_0^T \left(f(X(s))\right)^2 \mathrm{d}s\right] < \infty, \tag{2.7}$$

then $M_T := \int_0^T f(X(s))\mathrm{d}W(s)$ is a martingale, i.e.,

$$\mathbb{E}\left[\int_0^T f(X(s))\mathrm{d}W(s)\right] = 0.$$

2.4 Solution concepts for stochastic differential equations

One distinguishes between two main solution concepts for stochastic differential equations. A stochastic process $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^n$ can either define a solution in the strong sense and/or in the weak sense. In both cases X is required to solve the integral equation (2.5) almost surely, i.e.,

$$X(\omega, t) - X(\omega, 0) = \int_0^t \boldsymbol{a}(X(\omega, s))\mathrm{d}s + \int_0^t \boldsymbol{b}(X(\omega, s))\mathrm{d}W(\omega, s),$$

for almost all $\omega \in \Omega$. The difference between the two concepts lies in the way the in which the pair (X, W) is constructed which results in different restrictions on the probability space on which the Wiener process W and the solution X are defined and on the filtration $\mathcal{F} = (\mathcal{F}_t)_{t>0}$ to which the solution X is adapted. If, for a *given* Brownian motion W generating the filtration $\mathcal{F}$, there exists a stochastic process X, which is adapted to $\mathcal{F}$ and solves (2.5) almost surely, then X is referred to as a *strong solution* of the SDE (2.4). A weak solution is any pair (X, W), where W is a Wiener process generating the filtration $\mathcal{F}$ and X a stochastic process adapted to $\mathcal{F}$ so that (X, W) solves (2.5) almost surely. It follows directly from the definition that if for a given Wiener process W, the process X is a strong solution, then (X, W) is also a weak solution. Conversely, the existence of a weak solution of an SDE does in general not imply the existence of a strong solution. Tanaka's equation,

$$\mathrm{d}X = \mathrm{sign}(X)\mathrm{d}W$$

is a famous example for an SDE for which only weak solutions but no strong solutions exist. For details see e.g. [95, 51]. Uniqueness of the solution of the SDE (2.4) is to be understood in different ways depending on which solution concept is considered. Given an initial value $x_0 \in \mathbb{R}^n$, the SDE (2.4) is said to have a unique strong solution if for any two strong solutions $\widetilde{X}, X$ of (2.4) satisfying $\widetilde{X}(0) = X(0) = x_0$ it follows that

$$\widetilde{X}(\omega, t) = X(\omega, t), t \in [0, T], \text{ for almost all } \omega.$$

In the case of weak solutions this concept of uniqueness however is ill-defined, since infinitely many Wiener processes and thus potentially also infinitely many weak solutions (X, W) can exist on the same probability space. Instead, one distinguishes between two different notions of uniqueness for weak solutions.

Definition 2.4.1 (Pathwise uniqueness of weak solutions)**.** *Let $(\widetilde{X}, \widetilde{W})$ and (X, W) be weak solutions of the SDE (2.4) where $\widetilde{X}$ and X are adapted to the filtrations $\widetilde{\mathcal{F}}$ and $\mathcal{F}$, respectively and are defined on the same probability space so that $\widetilde{X}(0) = X(0)$. If $\widetilde{W} = W$ implies $\widetilde{X}(\omega, \cdot) = X(\omega, \cdot)$ for almost all ω, then it is said that pathwise uniqueness holds for solutions of the SDE (2.4).*

Definition 2.4.2 (Uniqueness in law of weak solutions)**.** *Uniqueness for the SDE (2.4) is said to hold in law if for any two weak solutions $\widetilde{X}, X$ with $\mathbb{P}(\widetilde{X}(0) = X(0)) = 1$, equality in law, i.e.,*

$$\mathbb{P}(X(t_i) \in B_i, i = 1, \ldots, k) = \mathbb{P}(\widetilde{X}(t_i) \in B_i, i = 1, \ldots, k),$$

holds for any $t_i > 0$ and Borel sets $B_i \subset \mathbb{R}^n, i = 1, \ldots, k$ with $k \in \mathbb{N}$.

It is easy to see from the definition that uniqueness in law does not imply pathwise

uniqueness nor uniqueness of strong solutions (provided a strong solution exists in the first place). Conversely, it directly follows from the definitions that the existence and uniqueness of a strong solution also implies pathwise uniqueness of a weak solution. Moreover, the following proposition can be shown to hold (for a proof see e.g. [51])

Proposition 2.4.1. *Pathwise uniqueness implies uniqueness in law.*

Proposition 2.4.2 (Uniqueness criteria for strong solutions, [95])**.** *Let $T > 0$. Assume*

$$|\boldsymbol{a}(x)| + |\boldsymbol{b}(x)| \leq C(1 + |x|),\ x \in \mathbb{R}^n,^1$$

some constant C and

$$|\boldsymbol{a}(x) - \boldsymbol{a}(y)| + |\boldsymbol{b}(x) - \boldsymbol{b}(y)| \leq D|x - y|,\ x, y \in \mathbb{R}^n,$$

for some constant D. Furthermore, let X_0 be a square integrable random variable, i.e.,

$$\mathbb{E}\left[|X_0|^2\right] < \infty,$$

which is independent of the Wiener process W, then the SDE (2.4) has a unique t-continuous solution $X(t), t \in [0, T]$ and

$$\mathbb{E}\left[\int_0^T |X(t)|^2 \mathrm{d}t\right] < \infty.$$

2.5 PDE description of weak solutions

For Markov processes which are the solution of an SDE, the action of the evolution operator $\mathcal{P}_t$ can be described by the solution of a partial differential equation. More specifically, let $u(x,t) := \mathcal{P}_t\varphi(x), t \geq 0$, then u is a solution of the Kolmogorov backward equation

$$\begin{aligned} \partial_t u(x,t) &= \mathcal{L}u(x,t), \\ u(x,0) &= \varphi(x), \end{aligned} \tag{2.8}$$

where the operator $\mathcal{L}$ is formally defined such that

$$(\mathcal{L}\varphi)(\boldsymbol{x}) = \lim_{\tau \to 0} \frac{(\mathcal{P}_t\varphi)(\boldsymbol{x}) - \varphi(\boldsymbol{x})}{\tau}, \tag{2.9}$$

for all $\varphi \in \mathscr{S}$. The operator $\mathcal{L}$ is commonly referred to as the *generator* associated with the SDE (2.4). If $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$, it takes the form of a differential operator, i.e.,

$$\mathcal{L} = -\boldsymbol{a} \cdot \nabla + \frac{1}{2}\boldsymbol{b}^\top \boldsymbol{b} \cdot \nabla^2. \tag{2.10}$$

This result is a direct consequence of Itô's lemma as shown in the following result:

Lemma 2.5.1. *Let $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$ be a compactly supported testfunction, and assume a weak solution of (2.4) exists, then*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}^x\left[\varphi(X(t))\right]\big|_{t=0} = \mathcal{L}\varphi(x), x \in \mathbb{R}^N$$

[1] Here and below in this proposition $|y|$ is to be interpreted as the Euclidian norm for $y \in \mathbb{R}^n$ and as the Frobenius norm for $y \in \mathbb{R}^{n \times n}$.

Proof. By Itô's formula

$$\varphi(X(t)) - \varphi(x) = \int_0^t \mathcal{L}\varphi(X_s)\mathrm{d}s + \int_0^t \nabla f^\top(X(s))\boldsymbol{b}(X(s))\mathrm{d}W(s), \tag{2.11}$$

the test function φ being smooth and having compact support, implies that the conditions of Proposition 2.3.1 are satisfied, hence

$$\mathbb{E}\left[\int_0^t \nabla f^\top(X(s))\boldsymbol{b}(X)\mathrm{d}W(s)\right] = 0.$$

Therefore, taking expectations in (2.11) yields,

$$\mathbb{E}^x\left[\varphi(X(t))\right] = \varphi(x) + \int_0^t \mathbb{E}^x \mathcal{L}\varphi(X_s)\mathrm{d}s \tag{2.12}$$

and thus

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}^x\left[\varphi(X_t)\right]\big|_{t=0} &= \lim_{t\to 0}\frac{\mathbb{E}^x\left[\varphi(X(t))\right] - \varphi(x)}{t} && (2.13)\\
&= \lim_{t\to 0}\frac{1}{t}\int_0^t \mathbb{E}^x\left[\mathcal{L}\varphi(X(s))\right]\mathrm{d}s && (2.14)\\
&= \mathcal{L}\varphi(x). && (2.15)
\end{aligned}$$

□

Kolmogorov forward equation As for the semigroup $\mathcal{P}_t, t > 0$ of operators acting on the test functions in $\mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$, the actions of operators of the adjoint semigroup $\mathcal{P}_t^\dagger, t > 0$ acting on the law of solution of the SDE (2.4), can be expressed as solutions of a PDE. Assume that the law of $X(t)$ has a density $\rho(\,\cdot\,, t) \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$ with respect to the Lebesque measure for $t > 0$, i.e.,

$$\rho(x,t)\mathrm{d}x = \mathbb{P}_{X(t)}(\mathrm{d}x).$$

Then by Lemma 2.5.1

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathbb{R}^n}\varphi(x)\rho(\mathrm{d}x,t)\right) &= \mathcal{L}\mathcal{P}_t\varphi(x)\\
&= \int_{\mathbb{R}^n}\mathcal{L}\varphi(x)\mathbb{P}_{X(t)}(\mathrm{d}x) && (2.16)\\
&= \int_{\mathbb{R}^n}\mathcal{L}\varphi(x)\rho(\mathrm{d}x,t)\mathrm{d}x.
\end{aligned}$$

is satisfied for all compactly supported $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$, which implies by the fundamental lemma of calculus of variations,

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(x,t) = \mathcal{L}^\dagger\rho(x,t), \tag{2.17}$$

where

$$\mathcal{L}^\dagger = \nabla\cdot(\boldsymbol{a}\,\cdot\,) + \frac{1}{2}\nabla^2 : (\boldsymbol{b}^T\boldsymbol{b}\,\cdot\,),$$

is the formal adjoint of $\mathcal{L}$ satisfying

$$\int_{\mathbb{R}^n} \mathcal{L}\varphi(x)\widetilde{\rho}(x)\mathrm{d}x = \int_{\mathbb{R}^n} \varphi(x)\mathcal{L}^\dagger\widetilde{\rho}(x)\mathrm{d}x, \tag{2.18}$$

for all $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$ and $\widetilde{\rho} \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$. The differential equation (2.17) is referred to as the Fokker-Planck equation.

Let ρ be a probability density. By (2.17) it follows that $\rho(x)\mathrm{d}x$ is an invariant probability measure for the SDE (2.4) if and only if

$$\mathcal{L}^\dagger\rho = 0. \tag{2.19}$$

Similarly, by (2.18) and the fundamental lemma of calculus of variation it follows that $\rho(x)\mathrm{d}x$ is an invariant density for the SDE (2.4) if and only if

$$\int_{\mathbb{R}^n} \mathcal{L}\varphi(x)\rho(x)\mathrm{d}x = 0, \tag{2.20}$$

for all compactly supported $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$.
Lastly, let π be a reference measure with a smooth density $\rho(x)\mathrm{d}x = \pi(x)$, then it follows by the same arguments as above (see e.g. [98]), that the action of the $L^2(\pi)$ adjoint semigroup $\mathcal{P}_t^*$, can be described by solutions of the PDE

$$\partial_t h(x,t) = \mathcal{L}^* h(x,t), \tag{2.21}$$

with

$$\mathcal{L}^* = \rho^{-1}\mathcal{L}^\dagger(\rho\,\cdot\,).$$

If a Markov processes is defined via an SDE, we use in the sequel the suggestive notation $e^{t\mathcal{L}}, e^{t\mathcal{L}^\dagger}$ and $e^{t\mathcal{L}^*}$ for the semigroups $\mathcal{P}_t, \mathcal{P}_t^\dagger$ and $\mathcal{P}_t^*$, respectively.

2.5.1 Extension to more general functional spaces

Up to this end we made the definition of the generator $\mathcal{L}$ precise for the case $\mathscr{S} = \mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$, by showing that under such regularity assumptions the generator takes the form of a differential operator as defined in (2.10). For more general choices of the test function set $\mathscr{S}$, the definition of $\mathcal{L}$ remains formal. The extension of $\mathcal{L}$ onto more general classes of test functions can be made precise in a functional analytic framework. For this purpose we consider the test function set $\mathscr{S}$ to be equipped with a norm $\|\cdot\|_{\mathscr{S}}$ such that $E := (\mathscr{S}, \|\cdot\|_{\mathscr{S}})$ forms a complete normed space, i.e., a Banach space. If $(\mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R}), \|\cdot\|_{\mathscr{S}})$ is a dense subspace of E, then the generator $\mathcal{L}$ can be continuously extended to the space E in a unique way. This is a direct consequence of the following Proposition 2.5.1, which can be found in any standard textbooks on functional analysis. (See e.g. [50, 105])

Proposition 2.5.1. *Let X be a normed linear space and Y a Banach space. If M is a dense linear subspace of X and*

$$T \,:\, M \subset X \to Y$$

is a bounded linear map, then there is a unique bounded linear map $\overline{T} : X \to Y$ such

that $\overline{T}x = Tx$ *for all* $x \in M$*. Moreover,*

$$\|T\|_{\mathcal{B}(M,Y)} = \|\overline{T}\|_{\mathcal{B}(X,Y)},$$

where for $(F, Z) \in \{(T, M), (\overline{T}, X)\}$ *the operator norm* $\|F\|_{\mathcal{B}(Z,Y)}$ *is defined as*

$$\|F\|_{\mathcal{B}(Z,Y)} = \sup_{x \in Z} \frac{\|Fx\|}{\|x\|}.$$

Examples of relevant functional spaces for which $\mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$ is a dense subset include

- the π-weighted L^2 (Hilbert-)space $\left(L^2(\pi), \|\cdot\|_{L^2(\pi)}\right)$, where $\|\cdot\|_{L^2(\pi)}$ denotes the norm induced by the π-weighted scalar product

  $$\langle f, g\rangle_{L^2(\pi)} = \int_{\mathbb{R}^n} f(x) \cdot g(x)\pi(\mathrm{d}x)$$

  and

  $$L^2(\pi) := \left\{\varphi \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}) : \|\varphi\|_{L^2(\pi)} < \infty\right\}.$$

- the π-weighted Sobolev space $\left(H^1(\pi), \|\cdot\|_{H^1(\pi)}\right)$, where $\|\cdot\|_{H^1(\pi)}$ denotes the norm induced by the scalar product

  $$\langle f, g\rangle_{H^1(\pi)} = \langle f, g\rangle_{L^2(\pi)} + \langle \nabla f, \nabla g\rangle_{L^2(\pi)}$$

  and

  $$H^1(\pi) := \left\{\varphi \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}) : \|\varphi\|_{H^1(\pi)} < \infty\right\} \subset L^2(\pi).$$

- weighted L^∞-spaces of the form $(L^\infty_{\mathcal{K}}, \|\cdot\|_{L^\infty_{\mathcal{K}}})$, where

  $$L^\infty_{\mathcal{K}} := \left\{\varphi \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}) : \frac{\varphi}{\mathcal{K}} \in L^\infty(\mathbb{R}^n)\right\}, \tag{2.22}$$

  and

  $$\|\varphi\|_{L^\infty_{\mathcal{K}}} := \left\|\frac{\varphi}{\mathcal{K}}\right\|_{L^\infty}, \tag{2.23}$$

  with $\mathcal{K} \in \mathcal{C}^2(\mathbb{R}^n, [1, \infty))$ satisfying the asymptotic growth condition

  $$\mathcal{K}(x) \to \infty \text{ as } \|x\| \to \infty.$$

While the extension $\overline{\mathcal{L}}$ of $\mathcal{L}$ to the above mentioned functional spaces can in general be not interpreted as a differential operator, certain properties of $\mathcal{L}$ are preserved. For example, it can be easily shown that the property (2.20) is inherited by $\overline{\mathcal{L}}$, i.e.,

$$\int_{\mathbb{R}^n} \overline{\mathcal{L}}\varphi(x)\pi(\mathrm{d}x) = 0, \tag{2.24}$$

for all $\varphi \in \mathscr{S}$, with $\mathscr{S} \in \left\{L^2(\pi), H^1(\pi), L^\infty_{\mathcal{K}}\right\}$. This is a simple consequence of the fact that

$$\mathbb{E}_\pi : g \mapsto \int_{\mathbb{R}^n} g(x)\pi(\mathrm{d}x).$$

considered as a (linear) operator on the above introduced functional spaces is bounded, thus continues. Similarly, under suitable assumption on the asymptotic properties of

the tails of the measure π, one can show that the action of the extensions of $\mathcal{L}$ when applied to $\mathcal{C}^\infty$-functions whose derivatives grow at most polynomially, coincides with the differential operator (2.10). In the remainder of this thesis we do not always clearly distinguish between $\mathcal{L}$ and its' extension $\overline{\mathcal{L}}$. Instead, following [76], we consider $\mathcal{L}$ in explicit calculations as an operator on a suitable dense functional subspace such that it can be interpreted as the differential operator (2.10).

Remark 2.5.1. *The derivation of the Kolmogorov backward equation in the proof of Lemma 2.5.1 requires the existence at least on a finite time interval, but not necessarily the uniqueness in law, of solutions of (2.4). Therefore, due to the lack of uniqueness, expectations in the proof of Lemma 2.5.1 are potentially ill defined. However, the limit (2.14) is still well defined as long as the expression $\mathbb{E}^x\mathcal{L}\varphi(X(t))$ is continuous in t. Consequently, the questions of uniqueness in law of weak solutions of the SDE (2.4) can be transferred to questions of uniqueness of the corresponding PDE.*
Conversely, if one can establish the existence and uniqueness of weak solutions of the SDE (2.4), (e.g. by requiring the coefficients $\boldsymbol{a}, \boldsymbol{b}$ to be sufficient regular so that the conditions of Proposition 2.4.2 are satisfied), then this implies that a unique solution of the Cauchy problem associated with (2.17) exists at least in a weak sense.

2.5.2 Hypoellipticity and existence of a smooth transition kernel.

Studying solutions of the Cauchy problem corresponding e.g. to (2.8) or studying solutions of the stationary Fokker-Planck equation is made significantly easier, if the solution is sufficiently regular. Within the scope of this thesis it is sufficient to consider the case where the differential operator $\partial_t - \mathcal{L}^\dagger$ is *hypoelliptic*. A differential operator A is said to be hypoelliptic, if for any g solving the differential equation $Ag = f$, it follows that g is of higher regularity than f in the sense that

$$f \in H_s^{\mathrm{loc}} \Rightarrow g \in H_{s+\epsilon}^{\mathrm{loc}},$$

with $\epsilon > 0$, where H_s^{loc} denotes the local Sobolev space of order $s \in \mathbb{N}$. This means that if $\partial_t - \mathcal{L}^\dagger$ is hypoelliptic, then the solution of (2.17) is smooth in the sense that $\rho \in \mathcal{C}^\infty(\mathbb{R}^n, [0,\infty))$ irrespective of the regularity of $\rho(\cdot, 0)$. A common way to establish hypoellipticity of a differential operator is via Hörmander's theorem ([48], Theorem 22.2.1, on page 353):

Theorem 2.5.1. *Let A be a differential operator of the form*

$$A = a_0 \cdot \nabla + \sum_{i=1}^{M} (a_i \cdot \nabla)^\dagger (a_i \cdot \nabla),$$

where $a_i, 0 \le i \le M$ are $\mathcal{C}^\infty$ vector fields in $\mathbb{R}^n$ and $\dagger$ indicates the formal L^2 adjoint. Iteratively define a collection of vector fields by

$$\mathscr{V}_0 = \{a_i : i \ge 0\}, \quad \mathscr{V}_{k+1} = \mathscr{V}_k \cup \{[\boldsymbol{v}, a_i] : \boldsymbol{v} \in \mathscr{V}_k,\, 0 \le i \le M\}, \tag{2.25}$$

where

$$[C, B] = (\nabla B)C - (\nabla C)B,$$

denotes the commutator of vector fields $C, B \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R}^n)$ and $(\nabla C), (\nabla B)$ the Jaco-

bian of the latters. If

$$\forall x \in \mathbb{R}^n, \quad \operatorname{lin}\left\{ \boldsymbol{v}(x) : \boldsymbol{v} \in \bigcup_{k\in\mathbb{N}} \mathscr{V}_k \right\} = \mathbb{R}^n, \tag{2.26}$$

then A is hypoelliptic.

The condition (2.26) is commonly referred to as *Hörmander's condition*. In the particular case of $A = \partial_t - \mathcal{L}^\dagger$ one can easily verify that (2.26) is exactly satisfied if

$$\mathscr{V}_0 = \{\boldsymbol{b}_i : i \geq 1\}, \quad \mathscr{V}_{k+1} = \mathscr{V}_k \cup \{[\boldsymbol{v}, \boldsymbol{b}_i] : \boldsymbol{v} \in \mathscr{V}_k, 0 \leq i \leq M\}.$$

with $\boldsymbol{b}_0 = \boldsymbol{a}$ and $\boldsymbol{b}_i$ refers to the i-th column of the diffusion coefficient $\boldsymbol{b}$ in (2.4). This particular version of Hörmander's condition adapted to the parabolic PDE of the form (2.17) is referred to as the *parabolic Hörmander condition*.

A simple calculation shows that $\mathcal{L}$ being hypoelliptic is equivalent to $\mathcal{L}^\dagger$ or $\mathcal{L}^*$ being hypoelliptic. Similarly, $\partial_t - \mathcal{L}$ being hypoelliptic is equivalent to $\partial_t - \mathcal{L}^\dagger$ or $\partial_t - \mathcal{L}^*$ being hypoelliptic.

A direct consequence of the parabolic Hörmander condition is that the transition kernel describing the evolution of the probability measure of solutions of the SDE possesses a smooth density, i.e., there are $p_t \in \mathcal{C}^\infty\left(\mathbb{R}^n \times \mathbb{R}^n, [0,\infty)\right),\ t > 0$, such that

$$\left(e^{t\mathcal{L}^\dagger}\mu_0\right)(\mathrm{d}x) = \int_{\mathbb{R}^n} p_t(y, \mathrm{d}x)\mu_0(\mathrm{d}y).$$

Similarly, hypoellipticity of $\mathcal{L}^\dagger$ implies that any invariant measure of the solution process of the associated SDE possesses a smooth density. However, it is important to note that hypoellipticity alone does not guarantee the existence of an invariant measure. The Brownian motion in $\mathbb{R}$ is a simple example of a process whose generator $\mathcal{L}$ is hypoellitpic (even $\partial_t - \mathcal{L}$ is hypoellitpic), but which does not possess an invariant measure.
Moreover, it can be shown (see [61]) that ergodicity (as defined in (2.1)) of the solution of an SDE follows if (i) there exists an invariant measure with positive smooth density and (ii) the associated generator $\mathcal{L}$ is hypoelliptic.

$$\begin{aligned} \dot{X} &= a, \\ X(0) &\sim \mu_0, \end{aligned}$$

with $a \in \mathbb{R}, a \neq 0$ and the domain of X is the one dimensional torus $\mathbb{T}$. It can be easily seen that the uniform distribution on $\mathbb{T}$ is the unique invariant measure of X. However, for $\mu_0 = \delta_x(\cdot)$, the law of $X(t)$ remains to be of the form of a Dirac-delta measure for all $t \geq 0$. In particular, the law of $X(t)$ does not converge to the uniform invariant distribution as $t \to \infty$.

2.5.3 Decay properties of semi-group operators and invertibility of the generator

Assume that the process $X(t)$ converges in law towards a unique invariant measure π, i.e.,

$$\lim_{t\to\infty} \mathbb{E}^x[\varphi(X(t))] = \mathbb{E}_\pi \varphi, \tag{2.27}$$

for all $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$, μ_0-almost all $x \in \mathbb{R}^n$. A common way to characterise the convergence of the expectation $\mathbb{E}^x[\varphi(X(t))]$ to the π-weighted average $\mathbb{E}_\pi\varphi$, or, more precisely the convergence of of $e^{t\mathcal{L}}\varphi(\,\cdot\,)$ to the constant function $x \mapsto \mathbb{E}_\pi\varphi$, is via functional decay estimates of the semi-group $e^{t\mathcal{L}}$. For this purpose let $E = (\mathscr{S}, \|\cdot\|_{\mathscr{S}})$ be a Banach space for which $\mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$ is a dense subset (see Section 2.5.1 for common choices).

Of particular interest in this context is exponential convergence of $e^{t\mathcal{L}}\varphi$ towards $\mathbb{E}_\pi\varphi$ in the respective norm, i.e.,

$$\|e^{t\mathcal{L}}\varphi - \mathbb{E}_\pi\varphi\|_{\mathscr{S}} \leq Ce^{-\kappa t}\|\varphi - \mathbb{E}_\pi\varphi\|_{\mathscr{S}}, \tag{2.28}$$

where C, κ are positive constants, the latter corresponding to the spectral gap of the generator $\mathcal{L}$ in the functional space $E_0 = (\mathscr{S}_0, \|\cdot\|_{\mathscr{S}})$, where $\mathscr{S}_0 \subseteq \mathscr{S}$ denotes the subset of test functions with vanishing mean, i.e.,

$$\mathscr{S}_0 = \{\varphi \in \mathscr{S} : \mathbb{E}_\pi\varphi = 0\}.$$

Let the operator Π denote the orthogonal projection from $\mathscr{S}$ onto $\mathscr{S}_0$, i.e.,

$$\Pi\varphi = \varphi - \mathbb{E}_\pi\varphi, \ \varphi \in \mathscr{S}.$$

Denote further by $\|\cdot\|_{\mathcal{B}(E)}$ the operator norm

$$\|\cdot\|_{\mathcal{B}(E)} := \sup_{\substack{\varphi \in E \\ \varphi \neq 0}} \frac{\|A\varphi\|_{\mathscr{S}}}{\|\varphi\|_{\mathscr{S}}},$$

of an operator $A : E \to E$. (2.28) implies that $e^{t\mathcal{L}}\Pi$ when considered as an operator on E is bounded in the operator norm $\mathcal{B}(E)$ as

$$\left\|e^{t\mathcal{L}}\Pi\right\|_{\mathcal{B}(E)} \leq Ce^{-\kappa t}. \tag{2.29}$$

Decay estimates of form (2.29) are of great importance as they allow to establish the invertibility of the generator $\mathcal{L}$ and the derivation of bounds for the inverse operator $\mathcal{L}^{-1}$ in $\mathcal{B}(E_0)$ as summarised in the following Proposition 2.5.2.

Proposition 2.5.2 ([76])**.** *Let E and E_0 be as specified above and assume that the estimate (2.29) holds. The generator $\mathcal{L}$ is invertible on E_0 and*

$$\|\mathcal{L}^{-1}\|_{\mathcal{B}(E_0)} \leq \frac{\kappa}{C} \tag{2.30}$$

with $C, \kappa > 0$ as in (2.29).

Following [76] the main steps of the proof of Proposition 2.5.2 can be summarised as follows. Let $D(\mathcal{L}) := \{\phi \in \mathscr{S}_0 : \mathcal{L}\phi \in \mathscr{S}_0\}$. Since

$$\mathcal{L}\left(-\int_0^\infty e^{t\mathcal{L}}\varphi \mathrm{d}t\right) = -\int_0^\infty \mathcal{L}e^{t\mathcal{L}}\varphi \mathrm{d}t = -\int_0^\infty \left(\frac{\mathrm{d}}{\mathrm{d}t}e^{t\mathcal{L}}\varphi\right)\mathrm{d}t = \varphi, \tag{2.31}$$

for $\varphi \in D(\mathcal{L})$, the inverse of $\mathcal{L}$ restricted to the domain $D(\mathcal{L})$ can be written as

$$\mathcal{L}^{-1} = -\int_0^\infty e^{t\mathcal{L}}\mathrm{d}t. \tag{2.32}$$

Moreover, since

$$\begin{aligned}\left\|\mathcal{L}^{-1}\Pi\varphi\right\|_{\mathscr{S}} &= \left\|\int_0^\infty e^{t\mathcal{L}}\Pi\varphi \mathrm{d}t\right\|_{\mathscr{S}} \\ &\leq \int_0^\infty \left\|e^{t\mathcal{L}}\Pi\varphi\right\|_{\mathscr{S}} \mathrm{d}t \\ &\leq C\int_0^\infty e^{-\kappa t}\|\Pi\varphi\|_{\mathscr{S}} \mathrm{d}t \\ &\leq \frac{C}{\kappa}\|\Pi\varphi\|_{\mathscr{S}},\end{aligned} \tag{2.33}$$

it follows that $\mathcal{L}^{-1}$ as an operator on $D(\mathcal{L})$ is bounded in the respective the respective operator norm.

Note that while $\varphi \in D(\mathcal{L})$ guarantees that each term in (2.31) is well defined, justifying the exchangeability of integration and the application of $\mathcal{L}$ would require more detailed arguments. Similarly, the extension to of the statment the whole space E_0 using the fact that $D(\mathcal{L})$ is a dense subspace of E_0 requires substantial work.

2.6 Geometric ergodicity

In this section we present results on the exponential convergence of the evolution operator associated with the semigroup $(\mathcal{P}_t)_{t\in\mathcal{T}}$. We concentrate on Lyapunov techniques which are particularly versatile as they are applicable both in the discrete time case $\mathcal{T} = \mathbb{N}$, as well as in the continuous time case $\mathcal{T} = [0, \infty)$. The convergence results derived via this approach hold in a suitable weighted L^∞ spaces, i.e., $E = (L^\infty_{\mathcal{K}}, \|\cdot\|_{L^\infty_{\mathcal{K}}})$, with $L^\infty_{\mathcal{K}}$ and $\|\cdot\|_{L^\infty_{\mathcal{K}}}$ as defined in (2.23) and (2.23), respectively. We note that there are several other frameworks, which allow the derivation of exponential decay estimates. In particular, we mention hypocoercivity techniques due to Villani [123], which for the case $\mathcal{T} = [0, \infty)$ allow the derivation of exponential decay estimates in $L^2_0(\pi) \cap H^1(\pi)$, where $L^2_0(\pi) \subset L^2(\pi)$ denotes the subspace of functions $\varphi \in L^2(\pi)$ with vanishing mean, i.e., $\mathbb{E}_\pi\varphi = 0$. Similarly, techniques proposed in [28] allow the derivation of decay estimates in $L^2_0(\pi)$.

The discrete time case

Let $X = (X_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ be a Markov chain with associated transition kernel $P(x, \mathrm{d}y)$. With some abuse of notation define

$$P\varphi(x) = \int_{\mathbb{R}^n} \varphi(y)P(x, \mathrm{d}y), \tag{2.34}$$

for any $\varphi \in L^\infty$.

Remark 2.6.1. *The evolution operator P as defined in (2.34) is identical to the element $\mathcal{P}_1$ of the semigroup $(\mathcal{P}_k)_{k\in\mathbb{N}}$. Also recall that the transition kernel $P(x, \mathrm{d}y)$ can be understood as the application of the semigroup operator $(\mathcal{P}^\dagger)_{k\in\mathbb{N}}$ on the Dirac measure δ_x, i.e.,*

$$P(x, \mathrm{d}y) = \mathcal{P}_1^\dagger\delta_x(\mathrm{d}y),$$

Moreover, if we assume that the transition kernel $P(x, \mathrm{d}y)$ possesses a smooth density, then it is also fully defined by expectations of indicator functions, i.e.,

$$P(x, S) = \mathbb{E}^x\left[\mathbf{1}_S(X_1)\right],$$

where S denotes a (Borel-)measurable subset of $\mathbb{R}^n$.

With the above notation we formulate the following assumptions:

Assumption 1 (Lyapunov condition)**.** *There exists a function $\mathcal{K} : \mathbb{R}^n \to [1, \infty)$ and constants $R \geq 0$ and $\alpha \in (0, 1)$ such that*

$$(P\mathcal{K})(x) \leq \alpha\mathcal{K}(x) + R, \text{ for all } x \in \mathbb{R}^n. \tag{2.35}$$

Assumption 2. *There exists a constant $\eta \in (0, 1)$ and a probability measure ν such that*

$$\inf_{x \in \mathcal{C}} P(x, \mathrm{d}y) \geq \eta\nu(\mathrm{d}y)$$

where $\mathcal{C} = \{x \in \mathbb{R}^n \,|\mathcal{K}(x) \leq \mathcal{K}_{\max}\}$ for some $\mathcal{K}_{\max} > 1 + 2R/(1 - \alpha)$, where α, R are the same constants as in Assumption 1.

It can be shown via a fixed point theorem that if Assumption 1 holds, i.e., if there exists a suitable Lyapunov function $\mathcal{K}$ satisfying Assumption 1, then this implies the existence of an ergodic measure for the Markov chain (See [12].) Assumption 2 ensures that the Markov chain is sufficiently rapidly mixing inside a compact set. Provided that this compact set is reachable from any point in phase space with positive probability this ensures that the Markov chain is irreducible, which in turn implies that the invariant measure (if existent) is unique. Together Assumption 1 and Assumption 2 ensure the ergodicity of the Markov chain X. This argumentation is made precise in the following theorem which is due to [45], and which also provides decay estimates for the semigroup $\mathcal{P}_n$. Similar results can be found in [12] and [86].

Theorem 2.6.1 ([45])**.** *Assume that Assumption 1 and Assumption 2 hold. Then P admits a unique invariant probability measure π so that $\mathcal{K} \in L^1(\pi)$, i.e.*

$$\int_{\Omega_{\boldsymbol{x}}} P(x, \mathrm{d}y)\pi(\mathrm{d}x) = \pi(\mathrm{d}y),$$

and

$$\int_{\Omega_{\boldsymbol{x}}} \mathcal{K}\mathrm{d}\pi < \infty. \tag{2.36}$$

Moreover, there exist $C > 0$ and $r \in (0, 1)$ such that for any $\varphi \in L^\infty_{\mathcal{K}}(\Omega_{\boldsymbol{x}})$ and any $n \in \mathbb{N}$,

$$\|P^n\varphi - \mathbb{E}_\pi\varphi\|_{L^\infty_{\mathcal{K}}} \leq Cr^n\|\varphi - \mathbb{E}_\pi\varphi\|_{L^\infty_{\mathcal{K}}}. \tag{2.37}$$

The continuous time case

Let X be the solution of the SDE (2.4) and let $\mathcal{L}$ denote the associated infinitesimal generator. The following two assumptions can be considered as continuous time analogues of Assumption 1 and Assumption 2, respectively.

Assumption 3 (infinitesimal Lyapunov condition)**.** *There is a function $\mathcal{K} \in \mathcal{C}^\infty(\Omega_{\boldsymbol{x}}, [1, \infty))$ with $\lim_{\|\boldsymbol{x}\|\to\infty} \mathcal{K}(x) = \infty$, and real numbers $a \in (0, \infty), b \in (0, \infty)$ such that,*

$$\mathcal{L}\mathcal{K} \leq -a\mathcal{K} + b. \tag{2.38}$$

Assumption 4. *For some $t > 0$ there exists a constant $\eta \in (0, 1)$ and a probability measure ν such that*

$$\inf_{x \in \mathcal{C}} e^{t\mathcal{L}^\dagger}\delta_x(\mathrm{d}y) \geq \eta\nu(\mathrm{d}y)$$

where $\mathcal{C} = \{x \in \Omega_{\boldsymbol{x}} : \mathcal{K}(x) \leq \mathcal{K}_{\max}\}$ *for some* $\mathcal{K}_{\max} > 1 + 2b/a$*, where* a, b *are the same constants as in (2.38).*

By integrating (2.38) over the time interval $[0, t']$, it directly follows that Assumption 3 implies Assumption 1 with $P = e^{t'\mathcal{L}}$. Similarly, if Assumption 4 is satisfied for $t = t'$ it is easy to see that this implies that Assumption 2 is satisfied for $P = e^{t'\mathcal{L}}$. Thus, if both (2.38) and Assumption 4 hold, once can then by virtue of Theorem 2.6.1 conclude geometric ergodicity for the embedded Markov chain associated with $P(x, \mathrm{d}y) = e^{t'\mathcal{L}^\dagger}\delta_x(\mathrm{d}y)$. The following theorem extends this result to the Markov process X itself:

Theorem 2.6.2 ([76])**.** *Let Assumption 3 hold and let there be* $t' > 0$ *so that Assumption 2 holds for* $P(x, \mathrm{d}y) = e^{t'\mathcal{L}^\dagger}\delta_x(\mathrm{d}y)$*. The solution of the SDE (2.4) admits a unique invariant probability measure* π *such that*

(i) for any $\varphi \in L^\infty_{\mathcal{K}}(\Omega_{\boldsymbol{x}})$ *we have*

$$\left\| e^{t\mathcal{L}}(\varphi) - \mathbb{E}_\pi \varphi \right\|_{L^\infty_{\mathcal{K}}} \leq \widetilde{C} e^{-t\lambda} \| \varphi - \mathbb{E}_\pi \varphi \|_{L^\infty_{\mathcal{K}}}, \tag{2.39}$$

where $\lambda > 0$ *such that* $e^{-t'\lambda} = r$*, and*

$$\widetilde{C} = C\left(1 + \frac{b}{a}\right) r^{-1},$$

with r *and* C *as in Theorem 2.6.1 and* a, b *as in Assumption 3.*

(ii)

$$\int_{\Omega_{\boldsymbol{x}}} \mathcal{K} \mathrm{d}\pi < \infty. \tag{2.40}$$

If, as in the case presented here, the stochastic process X is defined via an SDE, then a common way to show that a minorisation condition is satisfied, is via Lemma 2.6.1, which requires the following assumption to hold.

Assumption 5 ([81])**.** *Let* $X(t)$ *be a Markov process with associated semigroup* $(\mathcal{P}_t)_{t\geq 0}$ *and transition kernel* $P_t(x, dy)$ *and let* $C \subset \Omega_{\boldsymbol{x}}$ *a compact measurable set. If*

(i) for some $y^* \in int(C)$ *there is, for any* $\delta > 0$*, a* $t_1 = t_1(\delta) > 0$ *such that*

$$P_{t_1}(x, B_\delta(y^*)) > 0, \forall x \in C;$$

with $B_\delta(y^*) = \{x \in \Omega_{\boldsymbol{x}} : |x - y^*| < \delta\}$*.*

(ii) for any $t > 0$ *the transition kernel possesses a density* $p_t(x, y)$*, i.e.,*

$$(\mathcal{P}_t^\dagger \delta_x)(A) = \int_A p_t(x, y) dy, \ \forall x \in C, \ A \subset \Omega_{\boldsymbol{x}} \cap C, \ A \text{ measurable.}$$

and $p_t(x, y)$ *is jointly continuous in* $(x, y) \in C \times C$*.*

Lemma 2.6.1 ([81])**.** *Assumption 5* $\Rightarrow \exists\, t' > 0$ *so that*

(i) Assumption 2 holds for $P(x, \mathrm{d}y) = e^{t'\mathcal{L}^\dagger}\delta_x(\mathrm{d}y)$*.*

(ii) Assumption 4 holds for $t = t'$*.*

Assumption 6. *There is a $t_{\max} > 0$ so that for any $x^-, x^+ \in C$, there is a $t > 0$, with $t \leq t_{\max}$,so that the control problem*

$$\dot{\tilde{X}} = \boldsymbol{a}(\tilde{X}) + \boldsymbol{b}(\tilde{X})u, \tag{2.41}$$

subject to

$$\tilde{X}(0) = x^-, \text{ and } \tilde{X}(t) = x^+,$$

has a smooth solution $u \in \mathcal{C}^1([0, t_{\max}], \mathbb{R}^n)$.

Collorary 2.6.1. *If the SDE (2.4) satisfies Hörmander's condition and Assumption 6 holds, then Assumption 5 holds, which implies by virtue of Lemma 2.6.1 that also Assumption 2 holds.*

Girsanov's theorem provides conditions under which the path measures of two Itô processes are mutually absolutely continuous, which in particular implies that the law at any time $t \geq 0$ of these Itô process are equivalent. We will use Girsanov's theorem in Section 3.4 in order to proof the minorisation condition for instances of the GLE which in a Markovian representation possess coefficients which depend on the configurational variable. Here we provide a version of Girsanov's theorem which is adapted to Itô-diffusion processes.

Theorem 2.6.3 (Girsanov's theorem, [95])**.** *Let $\Omega_{\boldsymbol{x}} = \mathbb{T}^{n_1} \times \mathbb{R}^{n_2}, n = n_1 + n_2 \in \mathbb{N}$. Consider the two Itô diffusion processes*

$$\dot{X} = \boldsymbol{a}_x(X) + \boldsymbol{b}(X)\mathrm{d}\dot{\boldsymbol{W}};\ X(0) = x_0, \tag{2.42}$$

$$\dot{Y} = \boldsymbol{a}_y(Y) + \boldsymbol{b}(Y)\mathrm{d}\dot{\boldsymbol{W}};\ Y(0) = x_0, \tag{2.43}$$

where $x_0 \in \Omega_{\boldsymbol{x}}$, $\boldsymbol{W}$ is a standard Wiener process in $\mathbb{R}^n$, and $\boldsymbol{a}_x, \boldsymbol{a}_y : \Omega_{\boldsymbol{x}} \to \mathbb{R}^n$ and $\boldsymbol{b} : \Omega_{\boldsymbol{x}} \to \mathbb{R}^{n\times m}, m \in \mathbb{N}$, are such that there exist unique strong solutions X, Y for (2.42) *and* (2.43)*, respectively. If there is a function $\boldsymbol{u} \in \mathcal{C}(\Omega_{\boldsymbol{x}}, \mathbb{R}^n)$ such that*

$$\boldsymbol{a}_x - \boldsymbol{a}_y = \boldsymbol{b}\boldsymbol{u}$$

and $\boldsymbol{u}$ satisfies Novikov's *condition*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \|\boldsymbol{u}(X(t))\|_2^2 \mathrm{d}s\right)\right] < \infty. \tag{2.44}$$

then the path measures of X and Y on any finite time interval are equivalent. In particular, the support of the law of $X(t)$ and the support of the law of $Y(t)$ coincide for any $t > 0$.

2.7 Sampling with ergodic processes

Let X be the solution of the SDE (2.4). If X possesses a unique ergodic measure, then the Markov process can be used for sampling this measure. That is, the value of the integral

$$\mathbb{E}_\pi \varphi = \int_{\Omega_{\boldsymbol{x}}} \varphi(x)\pi(\mathrm{d}x),$$

is approximated by finite averages of the form

$$\bar{\varphi}_t := \frac{1}{t}\int_0^t \varphi(X(s))\mathrm{d}s.$$

The Monte-Carlo approximation $\bar{\varphi}_T$ is a stochastic quantity. We refer to the error in this approximation

$$\mathcal{E}_{\pi,t}[\varphi] := \bar{\varphi}_t - \mathbb{E}_\pi\varphi,$$

as the sampling error. The second moment of this quantity is referred to as the mean square error (MSE), i.e.,

$$\mathrm{MSE}(\pi,\varphi,t) := \mathbb{E}^x(\mathcal{E}_{\pi,t}[\varphi])^2.$$

It is common to consider a decomposition of the mean square error in terms of *bias* $\mu_t(\varphi)$ and variance $\sigma_t^2(\varphi)$ (see e.g. [31]), as

$$\mathbb{E}^x(\mathcal{E}_{\pi,t}[\varphi])^2 = (\mathbb{E}^x\bar{\varphi}_r - \mathbb{E}_\pi\varphi)^2 + \mathbb{E}^x\left(\bar{\varphi}_r - \mathbb{E}_\pi\varphi\right)^2 = (\mu_t(\varphi))^2 + \sigma_t^2(\varphi). \tag{2.45}$$

The bias $\mu_t(\varphi) = \mathbb{E}^x\bar{\varphi}_r - \mathbb{E}_\pi\varphi$ can be understood as the error due to an initialisation of the process out of equilibrium, i.e., when $X(0)$ is not initialised according to the invariant measure π. Asymptotically, this error can be shown (see [82]) to satisfy

$$(\mu_t(\varphi))^2 = O(t^{-1}).$$

and for sufficiently large $t > 0$ the variance can be shown to satisfy

$$\sigma_t^2(\varphi) \approx t^{-1}\xi_\varphi^2,$$

where ξ_φ^2 is the asymptotic variance of φ, which is defined as

$$\xi_\varphi^2 = \lim_{t\to\infty} t\,\mathbb{E}^x(\bar{\varphi}_t - \mathbb{E}_\pi\varphi)^2. \tag{2.46}$$

2.7.1 Central limit theorem for ergodic processes

Let $\varphi \in L^2(\pi)$. As shown in [15], a sufficient condition for the limit (2.46) to be well defined and for a central limit theorem of the form

$$\sqrt{t}\,\mathcal{E}_{\pi,t}[\varphi] \sim \mathcal{N}(0,\xi_\varphi^2), \text{ as } t\to\infty, \tag{2.47}$$

to hold, is that the Poisson equation

$$\mathcal{L}\Phi = \varphi - \mathbb{E}_\pi\varphi, \tag{2.48}$$

possesses a solution in $\Phi \in L^2(\pi)$. Note that for $\Phi \in L^2(\pi)$ it is a priori not clear how to interpret (2.48) since only under additional regularity assumptions (2.48) can be interpreted in a weak sense. A common way to makes sense of (2.48) is by deriving bounds for the operator $\mathcal{L}^{-1}$ in $\mathcal{B}(E_0)$ where E_0 is some subspace of $L_0^2(\pi)$ (see Section 2.5.3). If $\mathcal{L}^{-1}$ is bounded in $\mathcal{B}(E_0)$, this then directly implies $\Phi \in L^2(\pi)$ in (2.48) for $\varphi \in E_0$.

In the situation where decay estimates in spaces of the generic form $E = (L^\infty_{\mathcal{K}}, \|\cdot\|_{L^\infty_{\mathcal{K}}})$ are available, a central limit theorem for certain test functions can be derived as follows: Let $\mathcal{V}$ be a Lyapunov function such that the conditions for Theorem 2.6.2 are satisfied

for $\mathcal{K} = \mathcal{V}$. Note that if the conditions of Theorem 2.6.2 are also satisfied for $\mathcal{V}^2$, then this implies that a central limit theorem holds for all observables $\varphi \in L^\infty_{\mathcal{V}}$, since (2.40) being valid for $\mathcal{K} = \mathcal{V}^2$ implies

$$L^\infty_{\mathcal{V}} \subset L^2(\pi).$$

Thus, the inequality (2.33) for $\mathscr{S} = L^\infty_{\mathcal{V}^2}$ again implies that the solution Φ of (2.48) is contained in $L^2(\pi)$ for $\varphi \in L^\infty_{\mathcal{V}}$, so that by [15] indeed a central limit theorem of the form (2.47) holds for $\varphi \in L^\infty_{\mathcal{V}}$. This motivates to show the validity of Assumption 3 for a wide class of Lyapunov functions.

2.8 Stochastic splitting schemes

In this section we briefly revise a few basic results from stochastic numerical analysis regarding the error analysis of numerical weak approximation schemes for SDEs based on stochastic splittings to which we refer in the construction of stochastic numerical integrators for the GLE in later sections. For this purpose we consider the SDE (2.4) and assume a deterministic initial condition $X(0) = x_0$ as well as smooth coefficients $\boldsymbol{a} \in C^\infty(\mathbb{R}^n, \mathbb{R}^n)$ and $\boldsymbol{b} \in C^\infty(\mathbb{R}^n, \mathbb{R}^{n\times n})$. In the context of weak approximations we mean by a (stochastic) numerical integrator or numerical scheme, a map

$$\hat{\Phi}_{\Delta t} : \Omega_{\boldsymbol{x}} \times \Omega_{\mathscr{R}} \to \Omega_{\boldsymbol{x}}, \tag{2.49}$$

which together with a sequence of i.i.d random vectors $(\mathscr{R}^k)_{k\in\mathbb{N}} \subset \Omega_{\mathscr{R}}$ defines a Markov process $(\hat{X}_k)_{k\in\mathbb{N}}$ via the recursive relation

$$\hat{X}_{k+1} = \hat{\Phi}_{\Delta t}(\hat{X}_k, \mathscr{R}^k).$$

As we are only interested in weak approximations, we do not impose any relationship between the Wiener process W in the SDE and the sequence of random vectors $(\mathscr{R}^k)_{k\in\mathbb{N}}$ in the numerical integrator. For the sake of notational simplicity we will sometimes omit the random argument $\mathscr{R}^k$ in the integration map $\hat{\Phi}_{\Delta t}$ and write for example

$$\hat{X}_{k+1} = \hat{\Phi}_{\Delta t}(\hat{X}_k), \tag{2.50}$$

instead of (2.49). In what follows we describe how a stochastic integration map $\hat{\Phi}_{\Delta t}$ can be systematically constructed via a decomposition of the (stochastic) vector field corresponding to the right hand-side of SDE (2.4) and how the form of the semigroup $\hat{\mathcal{P}}^n_{\Delta t}$ associated with the discrete solution $(\hat{X}_k)_{k\in\mathbb{N}}$ can be derived in this case.

Stochastic flow map description

Let $\mathscr{R}$ be a random variable taking values in $\mathbb{R}^r$ with $r \in \mathbb{N}$ and consider a map $\Phi_{\Delta t} : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$. We refer to $\Phi_{\Delta t}(\,\cdot\,, \mathscr{R})$ as a *stochastic flow map* associated with the SDE (2.4), if

$$\mathbb{E}[\varphi(\Phi_{\Delta t}(x_0, \mathscr{R}))] = \mathbb{E}[\varphi(X(\Delta t)) \mid X(0) = x_0],$$

for all $x_0 \in \mathbb{R}^n$ and $\varphi \in \mathcal{C}^\infty_0(\mathbb{R}^n, \mathbb{R})$.

Consider a decomposition of (2.4) as

$$\boldsymbol{a}(X(t))\mathrm{d}t + \boldsymbol{b}(X(t))\mathrm{d}W(t) = \sum_{i=1}^{s} \boldsymbol{a}_{A_i}(X(t))\mathrm{d}t + \boldsymbol{b}_{A_i}(X(t))\mathrm{d}W_{A_i}(t), \tag{2.51}$$

such that $W_{A_i}, i = 1, \ldots, s$ are independent Wiener processess and the equality (2.51) holds in a weak sense. Let $\Phi_{\Delta t}^{A_i}$ denote the stochastic flow map associated with the SDE

$$\mathrm{d}X^{A_i} = \boldsymbol{a}_{A_i}(X^{A_i})\mathrm{d}t + \boldsymbol{b}_{A_i}(X^{A_i})\mathrm{d}W_{A_i}. \tag{2.52}$$

We refer to the (stochastic) map

$$\hat{\Phi}_{\Delta t}^{A_1,A_2,\ldots,A_s} = \Phi_{\Delta t}^{A_1} \circ \Phi_{\Delta t}^{A_2} \circ \cdots \circ \Phi_{\Delta t}^{A_s}, \text{[2]} \tag{2.53}$$

as a stochastic splitting scheme.

Semigroup description

By construction, the evolution operator

$$\hat{\mathcal{P}}_{\Delta t}^{A_1,\ldots,A_s}\varphi(x) := \mathbb{E}\left[\varphi\left(\hat{\Phi}_{\Delta t}^{A_1,A_2,\ldots,A_s}(x, \mathscr{R}^k)\right)\right], \text{ where } \varphi \in \mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R}),$$

which is associated with the stochastic splitting scheme (2.53), can be formally written as

$$\hat{\mathcal{P}}_{\Delta t}^{A_1,\ldots,A_s} = \exp(\Delta t \mathcal{L}_{A_s}) \exp(\Delta t \mathcal{L}_{A_{s-1}}) \cdots \exp(\Delta t \mathcal{L}_{A_1}),$$

where

$$\mathcal{L}_{A_i} = \boldsymbol{a}_{A_i} \cdot \nabla + \frac{1}{2}\boldsymbol{b}_{A_i} : \nabla^2,$$

denotes the infinitesimal generator associated with the SDEs (2.52) and $\exp(\Delta t \mathcal{L}_{A_s})$ is the corresponding evolution operator, i.e.,

$$\exp(t\mathcal{L}_{A_s})\varphi(x) = \mathbb{E}[\varphi(X^{A_i}(t)) \mid X^{A_i}(0) = x_0],$$

for $t \geq 0$.

Weak error analysis for stochastic splitting schemes

The weak error of a numerical SDE approximation refers to the discrepancy in the law of the numerical approximation and the exact solution process at finite time, or by duality to the difference between the respective expectation of an observable φ after finite time. More specifically, let $\hat{X}_0 = x_0$: The approximation of a numerical scheme is said to have weak order $p > 0$, if

$$|\mathbb{E}[\varphi(X(t_K))] - \mathbb{E}[\varphi(\hat{X}_K)]| \leq C(X_0, \varphi)\Delta t^p, \tag{2.54}$$

for fixed $t_K = K\Delta t \in [0, T]$ and sufficiently small Δt, holds for all testfunctions $\varphi \in \mathscr{S} \subset \mathcal{C}_0^{p+1}(\mathbb{R}^n, \mathbb{R})$, where $\mathcal{C}(X_0, \varphi) > 0$ is a constant which may depend on the initial condition x_0 and the testfunction φ. As in the deterministic case, the order of global weak convergence can be inferred from the convergence order of the weak error incurred

[2] For the sake of notational simplicity we suppress the random argument $\mathscr{R}$ here and in most below expressions.

in one step: Under some technical assumptions [87, Chapter 2] on the integrator map $\hat{\Phi}_{\Delta t}$, which ensure the existence of a stochastic Taylor expansion as we detail below, one can show that

$$|\mathbb{E}[\varphi(X(\Delta t))] - \mathbb{E}[\varphi(\hat{X}_1)]| \leq \tilde{C}(X_0, \varphi)\Delta t^{p+1}, \tag{2.55}$$

for sufficiently small step size $\Delta t > 0$, implies (2.54).

Taylor expansion of the evolution operator

As in the derivation of convergence orders of numerical schemes for ordinary differential equations, the weak order convergence of numerical schemes for stochastic differential equations can be determined via a Taylor expansion. The semigroup operator $\mathcal{P}_t$ can be formally expanded in the form of a Taylor series, i.e.,

$$\mathcal{P}_{\Delta t} = \sum_{k=0}^{\infty} \frac{\Delta t^k}{k!} \frac{\mathrm{d}^k \mathcal{P}_t}{\mathrm{d}t^k}\bigg|_{t=0}. \tag{2.56}$$

Let $\varphi \in \mathscr{S} \supset \mathcal{C}_0^\infty(\mathbb{R}^n, \mathbb{R})$ be an observable. By Taylor's theorem the remainder term $r_l(\Delta t, \varphi, x)$ for a truncation at order $l > 1$ is defined by the relation

$$(\mathcal{P}_{\Delta t}\varphi)(x) = \sum_{k=0}^{l} \frac{\Delta t^k}{k!} \frac{\mathrm{d}^k \mathcal{P}_t}{\mathrm{d}t^k}\bigg|_{t=0} \varphi(x) + r_l(\Delta t, \varphi, x), \tag{2.57}$$

and takes the form

$$r_l(\Delta t, \varphi, x) = \frac{\Delta t^{l+1}}{l!} \int_0^1 (1-\theta)^l \left(\frac{\mathrm{d}^{l+1} \mathcal{P}_t}{\mathrm{d}t^{l+1}} \right) \varphi(x)\bigg|_{t=\Delta t\theta} \mathrm{d}\theta \tag{2.58}$$

While the series (2.56) is in general not convergent [87], the expression (2.56) can be made explicit by suitably choosing the functional space $(\mathscr{S}, \|\cdot\|)$, so that the truncation error $r_l(\Delta t, \varphi, x)$ as a function of x is uniformly bounded for $\varphi \in \mathscr{S}$. In general it is however sufficient to ensure that the remainder (2.57) is bounded for every $\varphi \in \mathscr{S}$.

The Taylor expansion is used to derive weak convergence orders for numerical schemes in the following way: Let $e^{\Delta t \mathcal{L}}$ denote the semi group associated with an exact (weak) solution of the SDE (2.4), then according to (2.58):

$$(e^{\Delta t \mathcal{L}}\varphi)(x) = \sum_{k=0}^{l} \Delta t^k \mathcal{L}^k + r_l(\Delta t, \varphi, x), \tag{2.59}$$

with

$$r_l(\Delta t, \varphi, x) = \frac{\Delta t^{l+1}}{l!} \int_0^1 (1-\theta)^l \left(\mathcal{L}^{l+1} \exp(t\mathcal{L}) \right) \varphi(x)\bigg|_{t=\Delta t\theta} \mathrm{d}\theta. \tag{2.60}$$

Now considering the evolution operator associated with the splitting scheme $\hat{\mathcal{P}}_{\Delta t}^{A_1,\ldots,A_s}$, associated with the splitting scheme (2.53):

$$\hat{\mathcal{P}}_t^{A_1,\ldots,A_s} = e^{t\mathcal{L}_{A_s}} e^{t\mathcal{L}_{A_{s-1}}} \cdots e^{t\mathcal{L}_{A_1}} \tag{2.61}$$

The time derivative therefore takes the form

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\hat{\mathcal{P}}_t^{A_1,\ldots,A_s} =& \mathcal{L}_{A_s} e^{t\mathcal{L}_{A_s}} e^{t\mathcal{L}_{A_{s-1}}} \cdots e^{t\mathcal{L}_{A_1}} + e^{t\mathcal{L}_{A_s}} \mathcal{L}_{A_{s-1}} e^{t\mathcal{L}_{A_{s-1}}} \cdots e^{t\mathcal{L}_{A_1}} \\ &+ \cdots + e^{t\mathcal{L}_{A_s}} e^{t\mathcal{L}_{A_{s-1}}} \cdots \mathcal{L}_{A_1} e^{t\mathcal{L}_{A_1}},\end{aligned}$$

thus

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\hat{\mathcal{P}}_t^{A_1,\ldots,A_s}\right|_{t=0} = \mathcal{L}_{A_s} + \mathcal{L}_{A_{s-1}} + \cdots + \mathcal{L}_{A_1} = \mathcal{L}.$$

For higher derivatives it follows that

$$\left.\frac{\mathrm{d}^k}{\mathrm{d}t^k}\hat{\mathcal{P}}_t^{A_1,\ldots,A_s}\right|_{t=0} = \mathscr{T}^n\left[\mathcal{L}_{A_s},\cdots,\mathcal{L}_{A_1}\right] \tag{2.62}$$

where the expression

$$\mathscr{T}^n\left[\mathcal{L}_{A_s},\cdots,\mathcal{L}_{A_1}\right], \tag{2.63}$$

denotes the sum of all products where n operators of the set $\{\mathcal{L}_{A_s},\cdots,\mathcal{L}_{A_1}\}$ appear in the same order as they are listed in the rectangular bracket in expression (2.63).
It follows that

$$(\hat{\mathcal{P}}_{\Delta t}^{A_1,\ldots,A_s}\varphi)(x) \;\; = \sum_{k=0}^{l} \frac{\Delta t^k}{k!} \mathscr{T}^k\left[\mathcal{L}_{A_s},\cdots,\mathcal{L}_{A_1}\right]\varphi(x) + \hat{r}_l(\Delta t,\varphi,x). \tag{2.64}$$

If the remainder terms in (2.64) and (2.59) are bounded, then the convergence order follows simply by equating powers of Δt in (2.64) and (2.59).

2.8.1 Baker-Campbell-Hausdorff expansion

The algebraic form of the terms $\mathscr{T}^k\left[\mathcal{L}_{A_s},\cdots,\mathcal{L}_{A_1}\right]$, $k\in\mathbb{N}$ can be obtained by recursive application of the Baker-Campbell-Hausdorff (BCH) formula, which is given in terms of the formal operator identity

$$e^{\Delta t\mathcal{L}_{A_1}} e^{\Delta t\mathcal{L}_{A_2}} = e^{\Delta t\mathcal{A}}, \tag{2.65}$$

with

$$\begin{aligned}\mathcal{A} = (\mathcal{L}_{A_1} + \mathcal{L}_{A_2}) &+ \frac{\Delta t}{2}\left[\mathcal{L}_{A_1},\mathcal{L}_{A_2}\right] \\ &+ \frac{\Delta t^2}{12}\left(\left[\mathcal{L}_{A_1},\left[\mathcal{L}_{A_1},\mathcal{L}_{A_2}\right]\right] + \left[\mathcal{L}_{A_2},\left[\mathcal{L}_{A_2},\mathcal{L}_{A_1}\right]\right]\right) \\ &- \frac{\Delta t^3}{24}\left[\mathcal{L}_{A_2},\left[\mathcal{L}_{A_1},\left[\mathcal{L}_{A_1},\mathcal{L}_{A_2}\right]\right]\right] + O(\Delta t^4),\end{aligned} \tag{2.66}$$

where by

$$[A,B] = AB - BA,$$

we denote the commutator of the two linear operators A, B. Making use of the fact that $[A,B] = -[B,A]$ one easily derives

$$e^{\frac{\Delta t}{2}\mathcal{L}_{A_1}} e^{\Delta t\mathcal{L}_{A_2}} e^{\frac{\Delta t}{2}\mathcal{L}_{A_1}} = e^{\Delta t\mathcal{A}}. \tag{2.67}$$

with

$$\mathcal{A} = (\mathcal{L}_{A_1} + \mathcal{L}_{A_2}) + \frac{\Delta t^2}{12}\left([\mathcal{L}_{A_1}, [\mathcal{L}_{A_1}, \mathcal{L}_{A_2}]] - \frac{1}{2}[\mathcal{L}_{A_2}, [\mathcal{L}_{A_2}, \mathcal{L}_{A_1}]]\right) + O(\Delta t^5). \quad s \tag{2.68}$$

The splitting (2.65) is commonly referred to as a *Lie-Trotter splitting*. The symmetric splitting (2.67) is referred to as a *Strang splitting*. For splitting schemes of the general form $\hat{\mathcal{P}}_{\Delta t}^{A_1,\ldots,A_s}$ the terms $\mathscr{T}^k[\mathcal{L}_{A_s}, \cdots, \mathcal{L}_{A_1}]$, $k \in \mathbb{N}$ can be determined by recursively applying either (2.65),(2.66) and/or (2.67),(2.68).

Chapter 3

The quasi-Markovian generalised Langevin equation

In this thesis we focus on instances of the GLEs (1.28) and (3.25), which can be represented in an extended phase space as an Itô diffusion process. We refer to such GLEs as quasi-Markovian generalised Langevin equations (QGLE). In this chapter we introduce the general form of such SDE representations and discuss their asymptotic properties. The remainder of the chapter is organised as follows. In Section 3.1 and Section 3.2 we derive Markovian representations for the GLE (1.28) and (3.25), respectively. In Section 3.3 we briefly review results presented in the literature results on the Markovian representation and approximation of generalised Langevin equations. In Section 3.4 we derive results regarding the ergodic properties of the quasi-Markovian representations. In Section 3.5 we review results on the derivation of the underdamped Langevin equation and the overdamped Langevin equation as the limiting dynamics of the Markovian representation of (1.28) in certain asymptotic limits of the parametrisation of the Markovian representation.

3.1 Markovian representation of generalised Langevin equations with configuration independent noise

In this section we derive a Markovian respresentation of (1.28) starting with an SDE of the following form

$$\dot{\boldsymbol{q}} = \boldsymbol{M}^{-1}\boldsymbol{p}, \tag{3.1}$$

$$\dot{\boldsymbol{p}} = -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) - \boldsymbol{\Gamma}_{1,1}\boldsymbol{M}^{-1}\boldsymbol{p} - \boldsymbol{\Gamma}_{1,2}\boldsymbol{s} + \beta^{-1/2}\boldsymbol{\Sigma}_1\dot{\boldsymbol{W}}_t \tag{3.2}$$

$$\dot{\boldsymbol{s}} = -\boldsymbol{\Gamma}_{2,1}\boldsymbol{M}^{-1}\boldsymbol{p} - \boldsymbol{\Gamma}_{2,2}\boldsymbol{s} + \beta^{-1/2}\boldsymbol{\Sigma}_2\dot{\boldsymbol{W}}_t, \tag{3.3}$$

$$\text{with } (\boldsymbol{q}(0), \boldsymbol{p}(0), \boldsymbol{s}(0)) \sim \mu_0, \tag{3.4}$$

where $\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{M}$ are defined as above, $\boldsymbol{s}(t) \in \Omega_{\boldsymbol{s}} = \mathbb{R}^m$, and

$$\boldsymbol{\Gamma}_{1,1} \in \mathbb{R}^{n\times n}, \boldsymbol{\Gamma}_{1,2}, \boldsymbol{\Gamma}_{2,1}^T \in \mathbb{R}^{n\times m}, \boldsymbol{\Gamma}_{2,2} \in \mathbb{R}^{n\times n}, \boldsymbol{\Sigma}_1 \in \mathbb{R}^{n\times(n+m)}, \boldsymbol{\Sigma}_2 \in \mathbb{R}^{m\times(n+m)},$$

with $m \in \mathbb{N}$. The initial state of $(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s})$ is specified by the probability measure μ_0, which throughout this thesis we assume to be such that $(\boldsymbol{q}(0), \boldsymbol{p}(0), \boldsymbol{s}(0))$ has finite first

and second moments, in particular,

$$\int_{\Omega_{\boldsymbol{q}}\times\Omega_{\boldsymbol{s}}\times\Omega_{\boldsymbol{p}}} \|q\|_2^2 + \|p\|_2^2 + \|s\|_2^2 \; \mu_0(\mathrm{d}q, \mathrm{d}p, \mathrm{d}s) < \infty.$$

Notation: Throughout the remainder of this thesis we use the following shorthand notation:

$$\boldsymbol{\Gamma} := \begin{bmatrix} \boldsymbol{\Gamma}_{1,1} & \boldsymbol{\Gamma}_{1,2} \\ \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2} \end{bmatrix}, \qquad \boldsymbol{\Sigma} := \begin{bmatrix} \boldsymbol{\Sigma}_1 \\ \boldsymbol{\Sigma}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{1,1} & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{2,1} & \boldsymbol{\Sigma}_{2,2} \end{bmatrix}$$

With some abuse of notation we will also refer by $\boldsymbol{W}_1$ to the n first and by $\boldsymbol{W}_2$ the m last components of $\boldsymbol{W}$. Moreover, we use the following notation. We use $\boldsymbol{x}^T := (\boldsymbol{q}^T, \boldsymbol{p}^T, \boldsymbol{s}^T)$, as well as $\boldsymbol{z}^T := (\boldsymbol{p}^T, \boldsymbol{s}^T)$ as shorthand notations for the phase space variables. We use $\Omega_{\boldsymbol{x}} := \Omega_{\boldsymbol{q}} \times \Omega_{\boldsymbol{p}} \times \Omega_{\boldsymbol{s}}$, and $\Omega_{\boldsymbol{z}} := \Omega_{\boldsymbol{p}} \times \Omega_{\boldsymbol{s}}$, as shorthand notations for the corresponding domains. With some abuse of notation we also denote points in $\Omega_{\boldsymbol{x}}, \Omega_{\boldsymbol{z}}, \Omega_{\boldsymbol{q}}, \Omega_{\boldsymbol{p}}, \Omega_{\boldsymbol{s}}$ by $\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s}$, respectively.

Proposition 3.1.1. *Assume there exists a symmetric positive definite matrix $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ such that,*

$$\boldsymbol{\Gamma} \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix} + \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix} \boldsymbol{\Gamma}^T = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T, \tag{3.5}$$

then the SDE (3.1-3.3) conserves the probability measure $\mu_{\boldsymbol{Q},\beta}(d\boldsymbol{x})$ with density

$$\rho_{\boldsymbol{Q},\beta}(\boldsymbol{x}) \propto e^{-\beta[U(\boldsymbol{q})+\frac{1}{2}\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p}+\frac{1}{2}\boldsymbol{s}^T\boldsymbol{Q}^{-1}\boldsymbol{s}]}. \tag{3.6}$$

Conversely, assume that for a given symmetric positive definite matrix $\boldsymbol{Q}$ the SDE (3.1-3.3) conserves the probability measure $\mu_{\boldsymbol{Q},\beta}(d\boldsymbol{x})$, then $\boldsymbol{Q}$ satisfies the Lyapunov equation (3.5).

Proof. The statement follows by inspection of the stationary Fokker-Planck equation associated with the SDE (3.1-3.3). □

Proposition 3.1.2. *Let the condition of Proposition 3.1.1 be satisfied. If $-\boldsymbol{\Gamma}_{2,2}$ is a stable matrix, meaning that all eigenvalues of $\boldsymbol{\Gamma}_{2,2}$ have positive real parts, then the SDE (3.1-3.3) can be written in the form* (1.28) *with memory kernel $\boldsymbol{K} = \boldsymbol{K}_{\boldsymbol{\Gamma}}$, where*

$$\boldsymbol{K}_{\boldsymbol{\Gamma}}(t) \coloneqq \boldsymbol{\Gamma}_{1,1}\delta(t) - \boldsymbol{\Gamma}_{1,2}e^{-t\boldsymbol{\Gamma}_{2,2}}\boldsymbol{\Gamma}_{2,1}. \tag{3.7}$$

The coefficients of $\boldsymbol{K}_{\boldsymbol{\Gamma}}$ are integrable in the sense that

$$\int_0^\infty \boldsymbol{K}_{\boldsymbol{\Gamma},i,j}(t)\mathrm{d}t < \infty.$$

for $1 \le i, j \le n$.

Proof. The solution of (3.3) is found to be

$$\boldsymbol{s}(t) = e^{-\boldsymbol{\Gamma}_{2,2}t}\boldsymbol{s}(0) + \int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)}\boldsymbol{\Sigma}_2 d\boldsymbol{W}(s) - \int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)}\boldsymbol{\Gamma}_{2,1}\boldsymbol{M}^{-1}\boldsymbol{p}(s)ds.$$

Inserting this expression of $\boldsymbol{s}(t)$ into (3.2), we obtain

$$\begin{aligned}\dot{\boldsymbol{p}}(t) = &- \nabla_{\boldsymbol{q}} U(\boldsymbol{q}(t)) \\ &- \boldsymbol{\Gamma}_{1,1}\boldsymbol{M}^{-1}\boldsymbol{p}(t) + \boldsymbol{\Gamma}_{1,2}\int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)}\boldsymbol{\Gamma}_{2,1}\boldsymbol{M}^{-1}\boldsymbol{p}(s)ds \\ &+ \boldsymbol{\Sigma}_1 \mathrm{d}\boldsymbol{W}(t) - \boldsymbol{\Gamma}_{1,2}e^{-\boldsymbol{\Gamma}_{2,2}t}\boldsymbol{s}(0) - \boldsymbol{\Gamma}_{1,2}\int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)}\boldsymbol{\Sigma}_2 d\boldsymbol{W}(s)\end{aligned} \tag{3.8}$$

The second line in the above equation corresponds to the convolution term in the non-markovian formulation of the GLE. Now, by defining the random force as

$$\boldsymbol{\eta}(t) \coloneqq \boldsymbol{\Sigma}_1 \mathrm{d}\boldsymbol{W}(t) - \boldsymbol{\Gamma}_{1,2}e^{-\boldsymbol{\Gamma}_{2,2}t}\boldsymbol{s}(0) - \boldsymbol{\Gamma}_{1,2}\int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)}\boldsymbol{\Sigma}_2 d\boldsymbol{W}(s), \tag{3.9}$$

we obtain exactly the form of (1.28). Indeed, using the shorthand notation

$$\boldsymbol{G}(r) = \boldsymbol{\Gamma}_{1,2}\int_0^r e^{-\boldsymbol{\Gamma}_{2,2}(r-s)}\boldsymbol{\Sigma}_2 d\boldsymbol{W}(s).$$

Without loss of generality we assume that $t \geq t'$, and we find that the covariance of $\boldsymbol{\eta}$ has exactly the form (3.7):

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\eta}(t)\boldsymbol{\eta}^T(t')\right] &= \mathbb{E}\left[\boldsymbol{\Sigma}_1 \mathrm{d}\boldsymbol{W}(t)\mathrm{d}\boldsymbol{W}(t')^T\boldsymbol{\Sigma}_1^T\right] - \mathbb{E}[\boldsymbol{G}(t)\mathrm{d}\boldsymbol{W}^T(t')\boldsymbol{\Sigma}_1^T] \\ &\quad + \mathbb{E}\left[\boldsymbol{\Gamma}_{1,2}e^{-\boldsymbol{\Gamma}_{2,2}t}\boldsymbol{s}(0)\boldsymbol{s}(0)^T e^{-\boldsymbol{\Gamma}_{2,2}^T t'}\boldsymbol{\Gamma}_{1,2}^T\right] \\ &\quad + \mathbb{E}\left[\left(\boldsymbol{\Gamma}_{1,2}\int_0^{t'} e^{-\boldsymbol{\Gamma}_{2,2}(t-s)}\boldsymbol{\Sigma}_2 d\boldsymbol{W}(s)\right)\boldsymbol{G}^T(t')\right] \\ &= \delta(t-t')(\boldsymbol{\Gamma}_{1,1} + \boldsymbol{\Gamma}_{1,1}^T) - \boldsymbol{\Gamma}_{1,2}e^{-\boldsymbol{\Gamma}_{2,2}(t-t')}(\boldsymbol{\Gamma}_{2,1} + \boldsymbol{Q}\boldsymbol{\Gamma}_{1,2}^T) \\ &\quad + \boldsymbol{\Gamma}_{1,2}e^{-\boldsymbol{\Gamma}_{2,2}t}\boldsymbol{Q}e^{-\boldsymbol{\Gamma}_{2,2}^T t'}\boldsymbol{\Gamma}_{1,2}^T \\ &\quad + \boldsymbol{\Gamma}_{1,2}\int_0^{t'} e^{-\boldsymbol{\Gamma}_{2,2}(t-s)}(\boldsymbol{\Gamma}_{2,2}\boldsymbol{Q} + \boldsymbol{Q}\boldsymbol{\Gamma}_{2,2}^T)e^{-\boldsymbol{\Gamma}_{2,2}^T(t'-s)}\boldsymbol{\Gamma}_{1,2}^T ds \\ &= \delta(t-t')(\boldsymbol{\Gamma}_{1,1} + \boldsymbol{\Gamma}_{1,1}^T) - \boldsymbol{\Gamma}_{1,2}e^{-\boldsymbol{\Gamma}_{2,2}(t-t')}\boldsymbol{\Gamma}_{2,1},\end{aligned}$$

where expectations are taken over both μ_0 and the path measure of the Wiener process $\boldsymbol{W}$. The last equality follows by partial integration of the integral term. □

Provided that the conditions of Proposition 3.2.3 are satisfied, it follows that the projection on the position and momentum component of solutions of the SDE system (3.1-3.3) can be interpreted as weak solutions of the non-Markovian GLE (1.28) with $\boldsymbol{K} = \boldsymbol{K}_{\boldsymbol{\Gamma}}$.

3.1.1 Parametrisation of the extended variable form

In the previous subsection we have derived conditions for the SDE system (3.1-3.3) to correspond to the GLE (1.28). From a practical point of view it is typically of more interest to construct such an extended variable representation from a given memory kernel $\boldsymbol{K}$. For a general memory kernel $\boldsymbol{K}$, which might not be exactly representable in the extended variable formalism, there are various approaches which allow systematic approximation with respect to a certain predefined basis set corresponding to a particular structure of the drift matrix $\boldsymbol{\Gamma}$ (See e.g. [25, 88, 69].) We denote by M the set

of all triples $(\boldsymbol{\Gamma}, \boldsymbol{Q}, \boldsymbol{\Sigma})$ which satisfy the conditions of Proposition 3.2.3. Furthermore, we denote by $\widetilde{M}$ the set which contains all matrices $\boldsymbol{\Gamma}$ for which there exist suitable $\boldsymbol{Q}, \boldsymbol{\Sigma}$ so that $(\boldsymbol{\Gamma}, \boldsymbol{Q}, \boldsymbol{\Sigma}) \in M$. We say that a memory kernel $\boldsymbol{K}$ can be represented in quasi-Markovian form if there exists $\boldsymbol{\Gamma} \in \widetilde{M}$ so that $\boldsymbol{K} = \boldsymbol{K}_{\boldsymbol{\Gamma}}$. Given $(\boldsymbol{\Gamma}, \boldsymbol{Q}, \boldsymbol{\Sigma}) \in M$ it is straightforward to see that $\boldsymbol{\Gamma}, \boldsymbol{Q}$ uniquely determine $\boldsymbol{\Sigma}$. Conversely, it follows from Lemma A.0.2 that $-\boldsymbol{\Gamma}$ being stable is sufficient to ensure that the pair $\boldsymbol{\Gamma}, \boldsymbol{\Sigma}$ uniquely determines $\boldsymbol{Q}$.
We point out that the representation of the memory kernel in terms of $\boldsymbol{\Gamma}$ is not unique since the expression (3.7) is invariant under similarity transformations of the form

$$f_{\boldsymbol{S}} : \boldsymbol{\Gamma} \mapsto \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Gamma}_{1,1} & \boldsymbol{\Gamma}_{1,2} \\ \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}^{-1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma}_{1,1} & \boldsymbol{\Gamma}_{1,2}\boldsymbol{S}^{-1} \\ \boldsymbol{S}\boldsymbol{\Gamma}_{2,1} & \boldsymbol{S}\boldsymbol{\Gamma}_{2,2}\boldsymbol{S}^{-1} \end{pmatrix},$$

where $\boldsymbol{S} \in \mathbb{R}^{m\times m}$ is an invertible matrix. (See also [90], where the special case of $\boldsymbol{S}$ being an orthogonal matrix is mentioned).

Lemma 3.1.1. *Let $(\boldsymbol{\Gamma}, \boldsymbol{Q}, \boldsymbol{\Sigma}) \in M$, then, $\boldsymbol{\Gamma}_{1,1} = \boldsymbol{0}$ implies*

$$\boldsymbol{\Gamma}_{1,2}\boldsymbol{Q} = -\boldsymbol{\Gamma}_{2,1}^T. \tag{3.10}$$

Proof. Writing the left hand side of (3.5) in terms of the sub-blocks of $\boldsymbol{\Gamma}$ we find

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{\Gamma}_{1,2}\boldsymbol{Q} + \boldsymbol{\Gamma}_{2,1}^T \\ \boldsymbol{Q}\boldsymbol{\Gamma}_{1,2}^T + \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2}\boldsymbol{Q} + \boldsymbol{Q}\boldsymbol{\Gamma}_{2,2}^T \end{pmatrix}. \tag{3.11}$$

By Lemma A.0.1 (iii) it follows that this matrix is positive semi-definite if and only if (3.10) holds. □

Lemma 3.1.2. *Let $\boldsymbol{\Gamma} \in \widetilde{M}$ and $\boldsymbol{\Gamma}_{1,1} = \boldsymbol{0}$. If $\boldsymbol{S} \in \mathbb{R}^{m\times m}$ orthogonal, then*

$$f_{\boldsymbol{S}}(\boldsymbol{\Gamma}) \in \widetilde{M}.$$

3.1.2 Structural properties of the Markovian representation

In analogy to the underdamped Langevin equation the vector field defining the GLE can be written as the sum of the divergence-free Hamiltonian vector field H and a stochastic vector field O corresponding to the fluctuation-dissipation part:

$$\begin{bmatrix} \dot{\boldsymbol{q}} \\ \dot{\boldsymbol{p}} \\ \dot{\boldsymbol{s}} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{M}^{-1}\boldsymbol{p} \\ -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \\ \boldsymbol{0} \end{bmatrix}}_{=:\mathrm{H}} + \underbrace{\begin{bmatrix} \boldsymbol{0} \\ -\boldsymbol{\Gamma}\begin{pmatrix} \boldsymbol{M}^{-1}\boldsymbol{p} \\ \boldsymbol{s} \end{pmatrix} + \beta^{-1/2}\boldsymbol{\Sigma}\dot{\boldsymbol{W}}_t \end{bmatrix}}_{=:\mathrm{O}}. \tag{3.12}$$

Similarly, the infinitesimal generator (Kolmogorov backward operator) $\mathcal{L}_{\mathrm{GLE}}$, of (3.1-3.3) can be written as the sum of the Liouville operator corresponding to the Hamiltonian vector field and the generator of an Ornstein-Uhlenbeck process, i.e., $\mathcal{L}_{\mathrm{GLE}} = \mathcal{L}_H + \mathcal{L}_O$, where

$$\mathcal{L}_H := -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \cdot \nabla_{\boldsymbol{p}} + \boldsymbol{M}^{-1}\boldsymbol{p} \cdot \nabla_{\boldsymbol{q}}, \tag{3.13}$$

and

$$\mathcal{L}_O := -\boldsymbol{\Gamma}\begin{pmatrix} \boldsymbol{M}^{-1}\boldsymbol{p} \\ \boldsymbol{s} \end{pmatrix} \cdot \nabla_{\boldsymbol{z}} + \frac{\beta^{-1}}{2}\nabla_{\boldsymbol{z}}^2 : \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T. \tag{3.14}$$

One easily checks that the invariant measure (3.6) is preserved under the vector fields H and O, i.e.,

$$\mathcal{L}_H^\dagger \rho_{\boldsymbol{Q},\beta} = 0, \quad \mathcal{L}_O^\dagger \rho_{\boldsymbol{Q},\beta} = 0,$$

where $\mathcal{L}_H^\dagger$ and $\mathcal{L}_O^\dagger$ denote the respective adjoint operator (Kolmogorov forward operator) of $\mathcal{L}_H$ and $\mathcal{L}_O$. This establishes a formal analogy of the GLE in the extended variable formulation with the underdamped Langevin equation for the case of a simple scalar friction coefficient $\gamma > 0$ or positive definite friction matrix. As we will see in Section 4.7, this structural similarity allows to easily construct numerical integrators for the generalised Langevin equation by adapting existing numerical methods for the underdamped Langevin equation. However, while in the case of the underdamped Langevin equation the generator $\mathcal{L}_{\gamma,O} = -\gamma \boldsymbol{M}^{-1}\boldsymbol{p} \cdot \nabla_{\boldsymbol{p}} + \frac{\gamma}{\beta}\Delta_{\boldsymbol{p}}$ of the fluctuation-dissipation part of the vector field is elliptic, this is generally not true for $\mathcal{L}_O$ in the case of the GLE. This is a consequence of the fact that the dissipation matrix $\boldsymbol{\Gamma}$ in (3.1-3.3) is only required to have eigenvalues with positive real parts instead of being symmetric positive definite. In particular, unlike in the underdamped case, $\mathcal{L}_O$ is in general not self-adjoint in $L^2(\mu_{\boldsymbol{Q},\beta})$. Instead, the following proposition, which is a generalisation of Proposition 8.3 in [98], holds:

Proposition 3.1.3. *The operator $\mathcal{L}_O$ can be written as*

$$\mathcal{L}_O = \mathcal{A} + \mathcal{S}, \tag{3.15}$$

where

$$\mathcal{A} = -\boldsymbol{\Gamma}_{\mathcal{A}}\boldsymbol{z} \cdot \nabla_{\boldsymbol{z}}, \quad \mathcal{S} = -\boldsymbol{\Gamma}_{\mathcal{S}}\boldsymbol{z} \cdot \nabla_{\boldsymbol{z}} + \frac{\beta^{-1}}{2}\nabla_{\boldsymbol{z}}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T \cdot \nabla_{\boldsymbol{z}}), \tag{3.16}$$

with

$$\boldsymbol{\Gamma}_{\mathcal{A}} = \frac{1}{2}(\boldsymbol{\Gamma}\tilde{\boldsymbol{Q}} - \tilde{\boldsymbol{Q}}\boldsymbol{\Gamma}^T)\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{Q}}^{-1}, \quad \boldsymbol{\Gamma}_{\mathcal{S}} = \frac{1}{2}(\boldsymbol{\Gamma}\tilde{\boldsymbol{Q}} + \tilde{\boldsymbol{Q}}\boldsymbol{\Gamma}^T)\tilde{\boldsymbol{M}}^{-1}\tilde{\boldsymbol{Q}}^{-1}, \tag{3.17}$$

and

$$\tilde{\boldsymbol{Q}} = \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix}, \ \tilde{\boldsymbol{M}} = \begin{pmatrix} \boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_m \end{pmatrix}.$$

The operators $\mathcal{A}$ and $\mathcal{S}$ are antisymmetric and symmetric in $L^2(\mu_{\boldsymbol{Q},\beta})$, respectively i.e. $\mathcal{A} = -\mathcal{A}^*$ and $\mathcal{S} = \mathcal{S}^*$.

The property of $\mathcal{S}$ being self-adjoint in $L^2(\mu_{\boldsymbol{Q},\beta})$, is equivalent to the diffusion process $\boldsymbol{z}_t^{\mathcal{S}}$ associated with $\mathcal{S}$ being reversible [98]. We therefore refer to the (stochastic) vector field associated with $\mathcal{S}$ and the (deterministic) vector field associated with $\mathcal{A}$ as the reversible and irreversible parts, respectively. It directly follows from the above Proposition 3.1.3 that we can decompose the generator $\mathcal{L}_{\mathrm{GLE}}$ of (3.1-3.3) into antisymmetric and symmetric parts

$$\mathcal{L}_{\mathrm{GLE}} = \tilde{\mathcal{A}} + \mathcal{S},$$

where $\tilde{\mathcal{A}} = \mathcal{L}_H + \mathcal{A}$ and $\mathcal{A}, \mathcal{S}$ are defined in Proposition 3.1.3. Also note that the map $\boldsymbol{z} \mapsto -\boldsymbol{\Gamma}_{\mathcal{A}}\boldsymbol{z}$ defines a Hamiltonian vector field with Hamiltonian $\boldsymbol{z} \mapsto \frac{1}{2}\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p} + \frac{1}{2}\boldsymbol{s}^T\boldsymbol{Q}^{-1}\boldsymbol{s}$ and structure matrix

$$\boldsymbol{J}_{\mathcal{A}} = (\boldsymbol{\Gamma}\tilde{\boldsymbol{Q}} - \tilde{\boldsymbol{Q}}\boldsymbol{\Gamma}^T) = \begin{pmatrix} \boldsymbol{\Gamma}_{1,1} - \boldsymbol{\Gamma}_{1,1}^T & \boldsymbol{\Gamma}_{1,2}\boldsymbol{Q} - \boldsymbol{\Gamma}_{2,1}^T \\ -\boldsymbol{Q}\boldsymbol{\Gamma}_{1,2}^T + \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{\Gamma}_{2,2}^T \end{pmatrix}.$$

and one easily verifies that the vector field associated with the operator $\tilde{\mathcal{A}}$, i.e.,

$$\boldsymbol{x} \mapsto \begin{pmatrix} \boldsymbol{M}^{-1}\boldsymbol{p} \\ -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \\ \boldsymbol{0} \end{pmatrix} - \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\Gamma}_{\mathcal{A},1,1}\boldsymbol{p} + \boldsymbol{\Gamma}_{\mathcal{A},1,2}\boldsymbol{s} \\ \boldsymbol{\Gamma}_{\mathcal{A},2,1}\boldsymbol{p} + \boldsymbol{\Gamma}_{\mathcal{A},2,2}\boldsymbol{s} \end{pmatrix}, \tag{3.18}$$

is a Hamiltonian vector field with associated symplectic structure matrix

$$\boldsymbol{J}_{\tilde{\mathcal{A}}} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_n & \boldsymbol{0} \\ -\boldsymbol{I}_n & \boldsymbol{\Gamma}_{1,1} - \boldsymbol{\Gamma}_{1,1}^T & \boldsymbol{\Gamma}_{1,2}\boldsymbol{Q} - \boldsymbol{\Gamma}_{2,1}^T \\ \boldsymbol{0} & -\boldsymbol{Q}\boldsymbol{\Gamma}_{1,2}^T + \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2}\boldsymbol{Q} - \boldsymbol{Q}\boldsymbol{\Gamma}_{2,2}^T \end{pmatrix}.$$

3.1.3 Separability of the random force

Let $\boldsymbol{\Gamma} \in \widetilde{M}$. The form of the memory kernel in (3.7) suggests a decomposition as

$$\boldsymbol{K}_{\boldsymbol{\Gamma}}(t) = \boldsymbol{K}_{\boldsymbol{\Gamma}_c}(t) + \boldsymbol{K}_{\boldsymbol{\Gamma}_w}(t),$$

where

$$\boldsymbol{K}_{\boldsymbol{\Gamma}_w}(t) = \boldsymbol{\Gamma}_{1,1}\delta(t),$$

and

$$\boldsymbol{K}_{\boldsymbol{\Gamma}_c}(t) = -\boldsymbol{\Gamma}_{1,2} e^{-t\boldsymbol{\Gamma}_{2,2}} \boldsymbol{\Gamma}_{2,1} \boldsymbol{M},$$

i.e.,

$$\boldsymbol{\Gamma}_w = \begin{pmatrix} \boldsymbol{\Gamma}_{1,1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}, \ \boldsymbol{\Gamma}_c = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{\Gamma}_{1,2} \\ \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2} \end{pmatrix}.$$

It is clear that

$$\boldsymbol{\Gamma} \in \widetilde{M} \Rightarrow \boldsymbol{\Gamma}_w \in \widetilde{M},$$

holds true. However, it is in general not correct to assume

$$\boldsymbol{\Gamma} \in \widetilde{M} \Rightarrow \boldsymbol{\Gamma}_c \in \widetilde{M}.$$

For example, let

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma & \sqrt{\frac{\lambda\gamma}{\tau}} \\ \sqrt{\frac{\lambda\gamma}{\tau}} & \frac{1}{\tau} \end{pmatrix}, \tag{3.19}$$

with associated memory kernel

$$\boldsymbol{K}_{\boldsymbol{\Gamma}}(t) = \gamma\delta(t) - \lambda\frac{\gamma}{\tau} e^{-t/\tau}. \tag{3.20}$$

This memory kernel has been proposed in[126] where it was used to construct a sampling method with enhanced sampling properties. For $\lambda \in [0,1), \gamma > 0$ and $\tau > 0$ one easily verifies that $\boldsymbol{Q} = 1$ and $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T = \boldsymbol{\Gamma} + \boldsymbol{\Gamma}^T$ solves (3.5), thus $\boldsymbol{\Gamma} \in \widetilde{M}$. However, one finds $\boldsymbol{\Gamma}_c \notin \widetilde{M}$ since the Lyapunov equation (3.5) does not possess a solution in this case. This directly follows from the fact that the eigenvalues of the matrix

$$\begin{pmatrix} 0 & a \\ a & b \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} + \begin{pmatrix} 0 & a \\ a & b \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix} = \begin{pmatrix} 0 & ca+a \\ ca+a & 2bc \end{pmatrix},$$

are

$$\lambda_1 = bc - \sqrt{a^2c^2 + 2a^2c + a^2 + b^2c^2},\ \lambda_2 = \sqrt{a^2c^2 + 2a^2c + a^2 + b^2c^2} + bc,$$

thus $\lambda_1 < 0$, for $a, b > 0$, $c \geq 0$.

The above discussion motivates the following definition.

Definition 3.1.1. *Let $\mathbf{\Gamma} \in \widetilde{M}$. If also*

$$\begin{pmatrix} \mathbf{0} & \mathbf{\Gamma}_{1,2} \\ \mathbf{\Gamma}_{2,1} & \mathbf{\Gamma}_{2,2} \end{pmatrix} \in \widetilde{M},$$

then we say that the random force is $\boldsymbol{\eta}$ is separable.

Consider the decomposition of the random force as $\boldsymbol{\eta} = \boldsymbol{\eta}_w + \boldsymbol{\eta}_c$, where

$$\boldsymbol{\eta}_w(t) \coloneqq \mathbf{\Sigma}_1 \mathrm{d}\boldsymbol{W}(t),\ \boldsymbol{\eta}_c(t) := -\mathbf{\Gamma}_{1,2} e^{-\mathbf{\Gamma}_{2,2} t} \boldsymbol{s}(0) - \mathbf{\Gamma}_{1,2} \int_0^t e^{-\mathbf{\Gamma}_{2,2}(t-s)} \mathbf{\Sigma}_2 d\boldsymbol{W}(s). \quad (3.21)$$

Separability of the noise process implies independence of $\boldsymbol{\eta}_w$ and $\boldsymbol{\eta}_c$.

Lemma 3.1.3. *Let $\mathbf{\Gamma} \in \widetilde{M}$ such that the random force $\boldsymbol{\eta}$ is separable. The processess $\boldsymbol{\eta}_w$ and $\boldsymbol{\eta}_c$ as defined above are independent.*

Proof. By definition

$$\begin{pmatrix} \mathbf{0} & \mathbf{\Gamma}_{1,2} \\ \mathbf{\Gamma}_{2,1} & \mathbf{\Gamma}_{2,2} \end{pmatrix} \in \widetilde{M},$$

thus $\mathbf{\Gamma}_{1,2} = -\boldsymbol{Q}\mathbf{\Gamma}_{2,1}$ by Lemma 3.1.1, hence $\mathbf{\Sigma}_{1,2}^T = \mathbf{\Sigma}_{2,1} = \mathbf{0}$, which directly implies $\mathbb{E}[\boldsymbol{\eta}_w(t)\boldsymbol{\eta}_c^T(s)] = \mathbf{0}$[1] for all $t, s \geq 0$ □

The above introduced concept of separability of the random force is of practical relevance when it comes to design of numerical integrators for the SDE (3.1-3.3). As we discuss in Chapter 4 non-separability of the random form restricts the possible choices of decompositions in the design of splitting schemes for the SDE (3.1-3.3).

3.2 Markovian representation of generalised Langevin equations with configuration dependent noise

A simple extension of the markovian reformulation (3.1-3.3) of the GLE follows if we assume that the dissipation term and the diffusion term in that representation are smooth functions of $\boldsymbol{q}$. More specifically we consider an Itô diffusion of the form

$$\begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{M}^{-1}\boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) - \widetilde{\mathbf{\Gamma}}_{1,1}(\boldsymbol{q})\boldsymbol{M}^{-1}\boldsymbol{p} - \widetilde{\mathbf{\Gamma}}_{1,2}(\boldsymbol{q})\boldsymbol{s} + \beta^{-1/2}\widetilde{\mathbf{\Sigma}}_1(\boldsymbol{q})\dot{\boldsymbol{W}}, \\ \dot{\boldsymbol{s}} &= -\widetilde{\mathbf{\Gamma}}_{2,1}(\boldsymbol{q})\boldsymbol{p} - \widetilde{\mathbf{\Gamma}}_{2,2}(\boldsymbol{q})\boldsymbol{s} + \beta^{-1/2}\widetilde{\mathbf{\Sigma}}_2(\boldsymbol{q})\dot{\boldsymbol{W}}, \\ &\text{with } \boldsymbol{x}(0) \sim \mu_0, \end{aligned} \quad (3.22)$$

[1] The expectation is taken with respect to the inital distribution μ_0 and the path measure of the Brownian motion $\boldsymbol{W}$.

where $\widetilde{\boldsymbol{\Gamma}}_{1,1} \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}}, \mathbb{R}^{n\times n})$, $\widetilde{\boldsymbol{\Gamma}}_{2,1}^T, \widetilde{\boldsymbol{\Gamma}}_{1,2} \in (\Omega_{\boldsymbol{q}}, \mathbb{R}^{n\times m})$, $\widetilde{\boldsymbol{\Gamma}}_{2,2} \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}}, \mathbb{R}^{m\times m})$, $\widetilde{\boldsymbol{\Sigma}}_1 \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}}, \mathbb{R}^{n\times(n+m)})$, $\widetilde{\boldsymbol{\Sigma}}_2 \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}}, \mathbb{R}^{m\times(n+m)})$. Furthermore we let $\mu_0, \boldsymbol{W}, \beta, U$ be as defined in the preceding section. In the remainder of this thesis we use the compact notation

$$\widetilde{\boldsymbol{\Gamma}} = \begin{pmatrix} \widetilde{\boldsymbol{\Gamma}}_{1,1} & \widetilde{\boldsymbol{\Gamma}}_{1,2} \\ \widetilde{\boldsymbol{\Gamma}}_{2,1} & \widetilde{\boldsymbol{\Gamma}}_{2,2} \end{pmatrix} \in \mathcal{C}^\infty\left(\Omega_{\boldsymbol{q}}, \mathbb{R}^{(n+m)\times(n+m)}\right).$$

and

$$\widetilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \widetilde{\boldsymbol{\Sigma}}_{1,1} & \widetilde{\boldsymbol{\Sigma}}_{1,2} \\ \widetilde{\boldsymbol{\Sigma}}_{2,1} & \widetilde{\boldsymbol{\Sigma}}_{2,2} \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{\Sigma}}_1 \\ \widetilde{\boldsymbol{\Sigma}}_2 \end{pmatrix} \in \mathcal{C}^\infty\left(\Omega_{\boldsymbol{q}}, \mathbb{R}^{(n+m)\times(n+m)}\right).$$

Moreover, we denote the generator of (3.22) by

$$\widetilde{\mathcal{L}}_{\mathrm{GLE}} = \mathcal{L}_H + \widetilde{\mathcal{L}}_O,$$

with

$$\mathcal{L}_H := -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \cdot \nabla_{\boldsymbol{p}} + \boldsymbol{M}^{-1}\boldsymbol{p} \cdot \nabla_{\boldsymbol{q}},$$

and

$$\widetilde{\mathcal{L}}_O := -\widetilde{\boldsymbol{\Gamma}} \begin{pmatrix} \boldsymbol{M}^{-1}\boldsymbol{p} \\ \boldsymbol{s} \end{pmatrix} \cdot \nabla_{\boldsymbol{z}} + \frac{\beta^{-1}}{2} \nabla_{\boldsymbol{z}}^2 : \widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Sigma}}^T.$$

Although we expect that the results presented in the remainder of this section hold under some additional assumptions on the potential U and the functions $\widetilde{\boldsymbol{\Gamma}}, \widetilde{\boldsymbol{\Sigma}}$, also on an unbounded domain, we concentrate here on the case where $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n, L > 0$. We find this a reasonable choice as we expect this variation of the quasi-Markvian GLE to be of interest in physical models where a periodic position domain is assumed anyway. We first present a result, which ensures that the solution of (3.22) is well defined and has finite moments at all times.

Proposition 3.2.1. *Consider* (3.22) *with* $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n, L > 0$. *For any* $T > 0$ *the exists a unique strong solution* $\boldsymbol{x}(t), t \in [0, T]$ *of* (3.22), *which is continuous in* t *and*

$$\mathbb{E}\left[\int_0^T \|\boldsymbol{x}(t)\|_2^2 \mathrm{d}t\right] < \infty.$$

Proof. All coefficients in (3.22) are either linear in $\boldsymbol{z}$ or continuous functions of $\boldsymbol{q}$. Since $\Omega_{\boldsymbol{q}}$ is assumed to be compact, this directly implies that all coefficients are Lipschitz continuous. Since the initial distribution μ_0 of $\boldsymbol{x}(0)$ is assumed to be such that

$$\int_{\Omega_{\boldsymbol{x}}} \|x\|_2^2 \mu_0(\mathrm{d}x) < \infty,$$

holds, the statement follows from Proposition 2.4.2. □

The following theorem can be viewed as a generalisation of Proposition 3.1.1.

Proposition 3.2.2. *Assume there exists a symmetric positive definite matrix* $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ *such that for all* $\boldsymbol{q} \in \Omega_{\boldsymbol{q}}$,

$$\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q}) \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix} + \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix} \widetilde{\boldsymbol{\Gamma}}^T(\boldsymbol{q}) = \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{q})\widetilde{\boldsymbol{\Sigma}}^T(\boldsymbol{q}), \tag{3.23}$$

then the SDE (3.22) *conserves the probability measure* $\mu_{\boldsymbol{Q},\beta}(\mathrm{d}x)$, *which takes the same*

form as in Proposition 3.2.2.

Proof. The statement follows again by inspection of the stationary Fokker-Planck equation associated with the SDE (3.22). □

We emphasize that while the coefficients $\widetilde{\boldsymbol{\Gamma}}, \widetilde{\boldsymbol{\Sigma}}$ are functions of $q \in \Omega_{\boldsymbol{q}}$, the covariance matrix $\boldsymbol{Q}$ of the auxiliary variable $\boldsymbol{s}$ must not depend on the value of $\boldsymbol{q}$ if $\mu_{\boldsymbol{Q},\beta}$ is to be conserved under the dynamics of (3.22).

In what follows we relate the system (3.22) back to a non-Markovian (stochastic) integro-differential equation. We do this for a slightly simplified form of (3.22). More specifically, we assume

$$\widetilde{\boldsymbol{\Sigma}}_{1,2}(q) = \widetilde{\boldsymbol{\Sigma}}_{2,1}^T(q) = \boldsymbol{0},$$

and

$$\widetilde{\boldsymbol{\Gamma}}(q) = \boldsymbol{\Gamma}_{2,2} \in \mathbb{R}^{m\times m},\ \widetilde{\boldsymbol{\Sigma}}(q) = \boldsymbol{\Sigma}_{2,2} \in \mathbb{R}^{m\times m},$$

for all $q \in \Omega_{\boldsymbol{q}}$. We introduce the following convolution functional as a generalisation of the convolution term in the GLE (1.28),

$$\widetilde{\boldsymbol{K}_{\widetilde{\boldsymbol{\Gamma}}}}(\boldsymbol{q},t) * \boldsymbol{p} = \widetilde{\boldsymbol{\Gamma}}_{1,1}(\boldsymbol{q}(t))\boldsymbol{M}^{-1}\boldsymbol{p}(t) - \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}(t)) \int_0^t e^{-(t-s)\boldsymbol{\Gamma}_{2,2}} \widetilde{\boldsymbol{\Gamma}}_{2,1}(\boldsymbol{q}(s))\boldsymbol{M}^{-1}\boldsymbol{p}(s)\mathrm{d}s.$$

Moreover, we consider a random force of the form

$$\widetilde{\boldsymbol{\eta}}(t) = \widetilde{\boldsymbol{\eta}}_w(t) + \widetilde{\boldsymbol{\eta}}_c(t),$$

where

$$\widetilde{\boldsymbol{\eta}}_w(t) = \beta^{-1/2}\widetilde{\boldsymbol{\Sigma}}_1(\boldsymbol{q}(t))\dot{\boldsymbol{W}}_1(t),$$

and

$$\widetilde{\boldsymbol{\eta}}_c(t) = -\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}(t))\boldsymbol{\eta}_c(t),$$

with $\boldsymbol{\eta}_c$ being the solution of the Ornstein-Uhlenbeck process

$$\dot{\boldsymbol{\eta}}_c = -\boldsymbol{\Gamma}_{2,2}\boldsymbol{\eta}_c + \boldsymbol{\Sigma}_{2,2}\dot{\boldsymbol{W}}_2, \quad \boldsymbol{\eta}_c(0) = \boldsymbol{s}(0), \tag{3.24}$$

where $\boldsymbol{s}(0)$ corresponds to the initial value of $\boldsymbol{s}$ in (3.22). As shown in the following proposition the SDE (3.22) can under this assumptions be rewritten in the form of the GLE (1.30) as

$$\begin{aligned} \dot{\boldsymbol{q}}(t) &= \boldsymbol{M}^{-1}\boldsymbol{p}(t), \\ \dot{\boldsymbol{p}}(t) &= -\nabla_{\boldsymbol{q}}U(\boldsymbol{q}(t)) - \widetilde{\boldsymbol{K}_{\widetilde{\boldsymbol{\Gamma}}}}(\boldsymbol{q},t) * \boldsymbol{p} + \widetilde{\boldsymbol{\eta}}(t). \end{aligned} \tag{3.25}$$

Proposition 3.2.3. *Let $\widetilde{\boldsymbol{\Sigma}}_{1,2}^T(q) = \widetilde{\boldsymbol{\Sigma}}_{2,1}(q) = \boldsymbol{0}$, $\widetilde{\boldsymbol{\Sigma}}_{2,2}(q) = \boldsymbol{\Sigma}_{2,2} \in \mathbb{R}^{m\times m}$ for all $q \in \Omega_{\boldsymbol{q}}$, and $\widetilde{\boldsymbol{\Gamma}}_{2,2}(q) = \boldsymbol{\Gamma}_{2,2} \in \mathbb{R}^{m\times m}$, such that $-\widetilde{\boldsymbol{\Gamma}}(q)$ is stable for all $q \in \Omega_{\boldsymbol{q}}$. The SDE (3.22) can be rewritten in the form (3.25).*

Proof. The solution for $\boldsymbol{s}$ in (3.22) can be written as

$$\boldsymbol{s}(t) = e^{-\boldsymbol{\Gamma}_{2,2}t}\boldsymbol{s}(0) + \int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)}\boldsymbol{\Sigma}_{2,2}d\boldsymbol{W}_2(s) - \int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)}\widetilde{\boldsymbol{\Gamma}}_{2,1}(\boldsymbol{q}(s))\boldsymbol{M}^{-1}\boldsymbol{p}(s)ds. \tag{3.26}$$

By substituting $\boldsymbol{s}(t)$ in the second equation of (3.22) by the right hand side of (3.26) we obtain

$$\begin{aligned}\dot{\boldsymbol{p}}(t) =& -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}(t)) - \widetilde{\boldsymbol{\Gamma}}_{1,1}(\boldsymbol{q}(t))\boldsymbol{p}(t) + \boldsymbol{\Gamma}_{1,2}(\boldsymbol{q}(t)) \int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)} \boldsymbol{\Gamma}_{2,1}(\boldsymbol{q}(s)) \boldsymbol{M}^{-1} \boldsymbol{p}(s) \mathrm{d}s \\ &+ \widetilde{\boldsymbol{\Sigma}}_{1,1} \mathrm{d}\boldsymbol{W}_1(t) - \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}(t)) e^{-\boldsymbol{\Gamma}_{2,2} t} \boldsymbol{s}(0) - \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}(t)) \int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)} \boldsymbol{\Sigma}_{2,2} \mathrm{d}\boldsymbol{W}_2(s).\end{aligned}$$

Noting that $\boldsymbol{\eta}_c(t)$ as the solution of (5.9) can be written as

$$\boldsymbol{\eta}_c(t) = e^{-\boldsymbol{\Gamma}_{2,2} t} \boldsymbol{s}(0) - \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}(t)) \int_0^t e^{-\boldsymbol{\Gamma}_{2,2}(t-s)} \boldsymbol{\Sigma}_{2,2} \mathrm{d}\boldsymbol{W}_2(s)$$

we find that

$$\begin{aligned}\dot{\boldsymbol{p}}(t) &= -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}(t)) - \widetilde{\boldsymbol{K}}_{\widetilde{\boldsymbol{\Gamma}}}(\boldsymbol{q}, t) * \boldsymbol{p} + \widetilde{\boldsymbol{\eta}}_w(t) - \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}(t)) \boldsymbol{\eta}_c(t) \\ &= -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}(t)) - \widetilde{\boldsymbol{K}}_{\widetilde{\boldsymbol{\Gamma}}}(\boldsymbol{q}, t) * \boldsymbol{p} + \widetilde{\boldsymbol{\eta}}(t).\end{aligned}$$

□

3.3 Markovian representations of the GLE in the literature

The Markovian representation (3.1-3.3) is of similar generality as the one presented in [22, 69] and the steps in the derivation are essentially the same (see also [98, Chapter 8]). Likewise, a derivation of a Markovian representation of the form (3.22) can for example be found in a slightly less general setup in [80]. Among what is presented in the above Sections 3.1 and 3.2, we believe that

- the concept of separability of the random force which we introduce in Section 3.1.3
- the explicit formulation of algebraic conditions (see Proposition 3.2.2) for the matrix functions $\widetilde{\boldsymbol{\Gamma}}$ and $\widetilde{\boldsymbol{\Sigma}}$, which allow to conclude the preservation of the measure $\mu_{\beta,\boldsymbol{Q}}$ under the dynamics of (3.22) and as such can be viewed as a restatement of the fluctuation dissipation theorem.

may be regarded as novel contributions.

We point out that besides the above mentioned generic frameworks, there are countless of Markovian representations of the GLE mentioned in the literature which are derived in the context of a particular physical model or application. For example, the Markovian representations of the GLE derived in [29, 2, 65, 101] can be considered as special instances of the SDE (3.1-3.3). Similarly, the non-equilibrium models studied in [35, 33, 34, 100, 101] can be represented in the form of (3.1-3.3). Markovian representations of the GLE with position dependent memory kernels, which can be viewed as instances of the SDE (3.22) can be found in [60, 93, 94, 79].

3.3.1 A sufficient condition for the existence of a Markovian representation

Let $\boldsymbol{\eta}$ be a real valued stationary Gaussian process with vanishing mean and covariance function $\boldsymbol{K} \in \mathcal{C}(\mathbb{R}, \mathbb{R})$, i.e.,

$$\forall s, t \in \mathbb{R},\ \mathbb{E}[\boldsymbol{\eta}(t)] = 0,\ \boldsymbol{K}(t) = \mathbb{E}[\boldsymbol{\eta}(s+t)\boldsymbol{\eta}(s)].$$

We denote by $\widehat{\mu}_{\boldsymbol{K}}$ the spectral measure of $\boldsymbol{K}$, i.e.,

$$\boldsymbol{K}(t) = \int_{\mathbb{R}} e^{ikt} \mathrm{d}\widehat{\mu}_{\boldsymbol{K}}(k).$$

Note that the existence of the spectral measure is a direct consequence of the following proposition, which is and adapted (and simplified) version of what is commonly referred to as Bochner's theorem.

Proposition 3.3.1. *A complex-valued function C with domain $\mathbb{R}$ is the covariance function of a continuous weakly stationary[2] random process on $\mathbb{R}^n$ with finite first and second moments, if and only if it can be represented as*

$$C(t) = \int_{\mathbb{R}} e^{itk} \mathrm{d}\mu(k)$$

where μ is a positive finite measure.

The above Proposition 3.3.1 is a simplified version of [114]. For a proof of the theorem we refer to any standard text book in Fourier analysis, such as [106, Chapter 1].

Assume that $\widehat{\rho}_{\boldsymbol{K}}$ possesses a density with respect to the Lebesgue measure, i.e.,

$$\widehat{\mu}_{\boldsymbol{K}}(\mathrm{d}k) = \widehat{\rho}_{\boldsymbol{K}}(k)\mathrm{d}k.$$

It has been observed in [100] (see also [35, 101] for similar results), that $(\widehat{\rho}_{\boldsymbol{K}}(k))^{-1}$ being of the form of a polynomial implies, that $\boldsymbol{\eta}$ can be rewritten as Markovian process in an extended phase space. This can be seen as a consequence of the following criteria for Markovianity:

Proposition 3.3.2. *If $p(k) = \sum_{m=1} c_m(-ik)^m$ is a polynomial with real coefficients and roots in upper half plane then the Gaussian process with spectral density $|p(k)|^{-2}$ is the solution of the stochastic differential equation*

$$p\left(-i\frac{\mathrm{d}}{\mathrm{d}t}\right)\boldsymbol{\eta}(t)\mathrm{d}t = \mathrm{d}\boldsymbol{W}(t)$$

The above proposition is quoted from [101]. A simple and self-contained proof is also provided in this reference. For a more comprehensive discussion, we refer to [32].

As detailed in [101] the inverse density $(\widehat{\rho}_{\boldsymbol{K}}(k))^{-1}$ being a polynomial indeed implies the applicability of Proposition 3.3.2: The measure $\widehat{\mu}_{\boldsymbol{K}}$ is by virtue of Bochner's

[2] A stochastic process $(X(t))_{t\in\mathbb{R}}$ with associated covariance function C is said to be weakly stationary if $\mathbb{E}[X(t)] = \mathbb{E}[X(t+s)] = 0$ and $C(0,s) = C(t,t+s)$ for all $t, s \in \mathbb{R}$. Since Gaussian processes are fully characterised by the mean and covariance function, a Gaussian processes is weakly stationary if and only if it is stationary.

theorem a positive measure. Therefore $\widehat{\rho}_{\boldsymbol{K}}$ must be a positive function, i.e., a positive polynomial of even degree, which in turn implies the existence of a suitable polynomial $p(k) = \sum_{m=1} c_m(-ik)^m$ with properties as stated in Proposition 3.3.2.

Proposition 3.3.2 has been used extensively in [35, 34, 100, 101]) to derive finite dimensional Markovian representations of heat bath models of the form (1.42). Similarly, Proposition 3.3.2 can be also used to derive suitable distributions for the spring constants and the heat bath particle masses in the Ford-Kac model which ensure that in the thermodynamic limit the solution trajectory of the distinguished particle converges weakly to the solution of a stochastic integro-differential equation which can be represented in a Markovian form; see [66, 65, 41].

3.4 Ergodic properties

In this section we provide criteria for geometric ergodicity for the Markovian representations of the GLE introduced in the previous section, i.e., we show under certain conditions that there exists a unique probability measure with smooth density $\mu(\mathrm{d}\boldsymbol{x}) = \rho(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$, such that

$$\exists\, \kappa > 0, C > 0,\ \forall\, \varphi \in L^\infty_{\mathcal{K}},\ \|\mu[\varphi] - e^{t\mathcal{L}}\varphi\|_{L^\infty_{\mathcal{K}}} \leq Ce^{-\kappa t}\|\mu[\varphi] - \varphi\|_{L^\infty_{\mathcal{K}}}, \tag{3.27}$$

and

$$\int_{\Omega_{\boldsymbol{x}}} \mathcal{K}(\boldsymbol{x})\mu(\mathrm{d}\boldsymbol{x}) < \infty, \tag{3.28}$$

where $\mathcal{L} \in \{\mathcal{L}_{\mathrm{GLE}}, \widetilde{\mathcal{L}}_{\mathrm{GLE}}\}$ and $\mathcal{K} \in \mathcal{C}^2\left(\Omega_{\boldsymbol{x}}[1,\infty)\right)$ is a suitable Lyapunov function. In particular, if the assumptions of Proposition 3.1.1 or Proposition 3.2.2 are satisfied, then

$$\mu(\mathrm{d}\boldsymbol{x}) = \mu_{\boldsymbol{Q},\beta}(\mathrm{d}\boldsymbol{x}),$$

where $\mu_{\boldsymbol{Q},\beta}$ is as defined in Proposition 3.1.1.

All results are derived using the techniques introduced in Section 2.6. That is, we show that (i) the minorisation condition (Assumption 4) is satisfied and (ii) a suitable Lyapunov function exists which satisfies Assumption 3 (or more generally the existence of a suitable class of Lyapunov functions of which each instance satisfies Assumption 3). We treat the cases $\Omega_{\boldsymbol{q}} = \mathbb{T}^n$ and $\Omega_{\boldsymbol{q}} = \mathbb{R}^n$ separately. In the situation $\Omega_{\boldsymbol{q}} = \mathbb{R}^n$, we show geometric ergodicity for constant coefficients $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$. For the case of a bounded domain $\Omega_{\boldsymbol{q}} = \mathbb{T}^n$ we can show geometric ergodicity for slightly more general forms of (3.1-3.3). For the case $\Omega_{\boldsymbol{q}} = \mathbb{T}^n$ we also show geometric ergodicity for the SDE (3.22).

3.4.1 Summary of main results

We first present results for the SDE (3.1-3.3). Let for the remainder of this subsection $\boldsymbol{\Gamma}, \boldsymbol{\Sigma}$ be such that

(i) $-\boldsymbol{\Gamma}$ is stable

(ii) the SDE (3.1-3.3) satisfies the parabolic Hörmander condition both in the presence of the force term ∇U and also for the case $U \equiv 0$. We provide algebraic conditions on $\boldsymbol{\Gamma}, \boldsymbol{\Sigma}$ which imply the parabolic Hörmander condition in Section 3.4.3.

(iii) the conditions of Proposition 3.1.1 are satisfied so that $\mu_{\boldsymbol{Q},\beta}$ is an invariant measure of (3.1-3.3) with $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ as defined in Proposition 3.1.1.

Theorem 3.4.1. *Let $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n, L > 0$. For any $l \in \mathbb{N}$ there exists $\mathcal{K}_l \in \mathcal{C}^\infty(L\mathbb{T}^n \times \mathbb{R}^{n+m})$ with*

$$\mathcal{K}_l(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = O(\|\boldsymbol{z}\|^{2l}), \quad \text{as } \|\boldsymbol{z}\| \to \infty,\ \boldsymbol{z} = \begin{pmatrix}\boldsymbol{p}\\ \boldsymbol{s}\end{pmatrix},$$

such that (3.27) *and* (3.28) *hold for $\mathcal{K} = \mathcal{K}_l$, $\mathcal{L} = \mathcal{L}_{\mathrm{GLE}}$ and $\mu = \mu_{\boldsymbol{Q},\beta}$.*

Proof. The validity of the minorisation condition follows from Lemma 3.4.2. The existence of a suitable class of Lyapunov functions is shown in Lemma 3.4.1. □

In the case of an unbounded configurational domain, i.e., $\Omega_{\boldsymbol{q}} = \mathbb{R}^n$, we require additional assumption on the potential function U in order to establish a suitable class of Lyapunov functions.

Assumption 7. *(i) the potential function is bounded from below, i.e., there exists $u_{\min} > -\infty$ such that*

$$\forall \boldsymbol{q} \in \Omega_{\boldsymbol{q}},\ U(\boldsymbol{q}) > u_{\min}.$$

(ii) there exist constants $D, E > 0$ and $F \in \mathbb{R}$ such that

$$\boldsymbol{q}^T \nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \geq DU(\boldsymbol{q}) + E\|\boldsymbol{q}\|_2^2 + F \text{ for all } \boldsymbol{q} \in \Omega_{\boldsymbol{q}}. \tag{3.29}$$

Theorem 3.4.2. *Let $\Omega_{\boldsymbol{q}} = \mathbb{R}^n$, U satisfies Assumption 7,* $\mathrm{rank}(\boldsymbol{\Sigma}) = n + m$ *and* $\mathrm{rank}(\boldsymbol{\Gamma}_{1,1}) = n$. *The probability measure $\mu_{\boldsymbol{Q},\beta}$ is the unique invariant measure of (3.1-3.3) and for any $l \in \mathbb{N}$ there exists $\mathcal{K}_l \in \mathcal{C}^\infty(\mathbb{R}^{2n+m}, [1,\infty))$ with*

$$\mathcal{K}_l(\boldsymbol{x}) = O(\|\boldsymbol{x}\|^{2l}), \quad \text{as } \|\boldsymbol{x}\| \to \infty,$$

such that (3.27) *and* (3.28) *hold for $\mathcal{K} = \mathcal{K}_l$, $\mathcal{L} = \mathcal{L}_{\mathrm{GLE}}$ and $\mu = \mu_{\boldsymbol{Q},\beta}$.*

Proof. The validity of a minorisation condition follows from Lemma 3.4.4. The existence of a suitable class of Lyapunov functions is shown in Lemma 3.4.3. □

The above theorem covers instances of the GLE with a non-degenerated white noise component. In order to derive geometric ergodicity for GLEs without a white noise component, i.e. $\boldsymbol{\Gamma}_{1,1} = \boldsymbol{0}$, we require U to be of the form of a perturbed quadratic potential function in the following sense:

Assumption 8. *Let the potential function U be such that*

$$U(\boldsymbol{q}) = U_1(\boldsymbol{q}) + U_2(\boldsymbol{q}),$$

where $U_1 \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ has bounded derivatives and

$$U_2(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T \boldsymbol{H} \boldsymbol{q},$$

with $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ is a positive definite matrix, i.e., $\min \sigma(\boldsymbol{H}) = \lambda_{\boldsymbol{H}} > 0$.

Remark 3.4.1. *Assumption 8 implies that there is $\overline{H} > 0$ and $\overline{h} \in \mathbb{R}$ so that*

$$|\langle \boldsymbol{g}, \nabla_{\boldsymbol{q}} U(\boldsymbol{q})\rangle| \leq \overline{H}|\langle \boldsymbol{g}, \boldsymbol{q}\rangle| + \overline{h},$$

for all $\boldsymbol{q}, \boldsymbol{g} \in \mathbb{R}^n$.

The following theorem provides condition for geometric ergodicity of (3.1-3.3) for the case $\boldsymbol{\Gamma}_{1,1} = \mathbf{0}$.

Theorem 3.4.3. *Let $\Omega_{\boldsymbol{q}} = \mathbb{R}^n$, U satisfies Assumption 7 and Assumption 8, and $\boldsymbol{\Gamma}_{1,1} = \mathbf{0}$. For any $l \in \mathbb{N}$ there exists $\mathcal{K}_l \in \mathcal{C}^\infty(\mathbb{R}^{2n+m}, [0,\infty))$ with*

$$\mathcal{K}_l(\boldsymbol{x}) = O(\|\boldsymbol{x}\|^{2l}), \quad \text{as } \|\boldsymbol{x}\| \to \infty,$$

such that (3.27) *and* (3.28) *hold for $\mathcal{K} = \mathcal{K}_l$, $\mathcal{L} = \mathcal{L}_{\mathrm{GLE}}$ and $\mu = \mu_{\boldsymbol{Q},\beta}$.*

Proof. The validity of the minorisation condition follows from Lemma 3.4.7. The existence of a suitable class of Lyapunov functions is shown in Lemma 3.4.3. □

For the case of a periodic configurational domain $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n, L > 0$ we show geometric ergodicity for the SDE (3.22). We focus on the case

$$\widetilde{\boldsymbol{\Gamma}}(\cdot) = \begin{pmatrix} \mathbf{0} & \widetilde{\boldsymbol{\Gamma}}_{1,2}(\cdot) \\ \widetilde{\boldsymbol{\Gamma}}_{2,1}(\cdot) & \widetilde{\boldsymbol{\Gamma}}_{2,2}(\cdot) \end{pmatrix} \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}}, \mathbb{R}^{2n\times 2n}).$$

where all non vanishing subblocks are assumed to be invertible, i.e.,

$$\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}), \widetilde{\boldsymbol{\Gamma}}_{2,1}(\boldsymbol{q}), \widetilde{\boldsymbol{\Gamma}}_{2,2}(\boldsymbol{q}), \widetilde{\boldsymbol{\Sigma}}_{2,2}(\boldsymbol{q}) \in \mathrm{GL}_n(\mathbb{R}),$$

for all $\boldsymbol{q} \in \Omega_{\boldsymbol{q}}$, where by $\mathrm{GL}_n(\mathbb{R}) \subset \mathbb{R}^{n\times n}$ we denote the set of all invertible $n \times n$-matrices with real valued coefficients. Furthermore, we assume that $-\widetilde{\boldsymbol{\Gamma}}(q)$ is a stable matrix for all $q \in \Omega_{\boldsymbol{q}}$ and that $\widetilde{\boldsymbol{\Gamma}}, \widetilde{\boldsymbol{\Sigma}}$ are such that the conditions of Proposition 3.2.2 are satisfied. These assumptions imply that $\mu_{\boldsymbol{Q},\beta}$ with $\boldsymbol{Q}$ as defined in the latter proposition is an invariant measure of (3.22) and the identity

$$\forall \boldsymbol{q} \in \Omega_{\boldsymbol{q}}, \ \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}) = -\boldsymbol{Q}\widetilde{\boldsymbol{\Gamma}}_{2,1}(\boldsymbol{q}), \tag{3.30}$$

holds. Moreover we assume

$$\exists\, \boldsymbol{C} \in \mathbb{R}^{(n+m)\times(n+m)} \text{ s.p.d. } \forall\, \boldsymbol{q} \in \Omega_{\boldsymbol{q}} : \ \widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q})\boldsymbol{C} + \boldsymbol{C}\widetilde{\boldsymbol{\Gamma}}^T(\boldsymbol{q}) \text{ s.p.d.}^3. \tag{3.31}$$

We expect that our result can be easily extended to more general forms of $\widetilde{\boldsymbol{\Gamma}}$ (see Remark 3.4.2). We also point out that the case $\widetilde{\boldsymbol{\Gamma}}_{1,1} \neq \mathbf{0}$ would not cause any additional difficulties in the proof of the result as long as the identity (3.30) holds. (See e.g. [108] for ergodicity results for underdamped Langevin equation with non-constant coefficients.)

Theorem 3.4.4. *Let $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n$, $L > 0$. Under the assumptions on $\widetilde{\boldsymbol{\Gamma}}$ and $\widetilde{\boldsymbol{\Sigma}}$ described in the preceding paragraph, there exists for any $l \in \mathbb{N}$ a function $\mathcal{K}_l \in \mathcal{C}^\infty(L\mathbb{T}^n \times \mathbb{R}^{2n}, [1,\infty))$ with*

$$\mathcal{K}_l(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = O(\|\boldsymbol{z}\|^{2l}), \quad \text{as } \|\boldsymbol{z}\| \to \infty, \ \boldsymbol{z} = \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{s} \end{pmatrix},$$

such that (3.27) *and* (3.28) *hold for $\mathcal{K} = \mathcal{K}_l$, $\mathcal{L} = \widetilde{\mathcal{L}}_{\mathrm{GLE}}$ and $\mu = \mu_{\boldsymbol{Q},\beta}$.*

Proof. The validity of the minorisation condition follows from Lemma 3.4.7. The existence of a suitable class of Lyapunov functions is shown in Lemma 3.4.10 □

[3] We use *s.p.d.* as the abbreviation for *symmetric positive definite*

Remark 3.4.2. *We expect that the results of Lemma 3.4.7 can be extended to instances of* (3.22)*, where* $\widetilde{\boldsymbol{\Gamma}}$ *is of a form such that in the non-markovian reformulation (3.25) the memory kernel is of the form*

$$\boldsymbol{K}_{\widetilde{\boldsymbol{\Gamma}}}(\boldsymbol{q},t) = \widetilde{\boldsymbol{\Gamma}}_{1,1}(\boldsymbol{q})\delta(t) - \sum_{i=1}^{K} \widetilde{\boldsymbol{\Gamma}}^{(i)}_{1,2}(\boldsymbol{q}) e^{-t\boldsymbol{\Gamma}^{(i)}_{2,2}} \widetilde{\boldsymbol{\Gamma}}^{(i)}_{2,1}(\boldsymbol{q}), \;\; K \in \mathbb{N},$$

where each $\widetilde{\boldsymbol{\Gamma}}^{(i)}$*,*

$$\widetilde{\boldsymbol{\Gamma}}^{(i)}(\boldsymbol{q}) = \begin{pmatrix} \mathbf{0} & \widetilde{\boldsymbol{\Gamma}}^{(i)}_{1,2}(\boldsymbol{q}) \\ \widetilde{\boldsymbol{\Gamma}}^{(i)}_{2,1}(\boldsymbol{q}) & \boldsymbol{\Gamma}^{(i)}_{2,2} \end{pmatrix}$$

satisfies the same conditions as $\widetilde{\boldsymbol{\Gamma}}$ *in Theorem 3.4.4.*

In order to demonstrate that the condition for Theorem 3.4.4 on $\boldsymbol{\Gamma}, \boldsymbol{\Sigma}$ are not too restrictive in the sense that the set of pairs of matrices $\boldsymbol{\Gamma}, \boldsymbol{\Sigma}$ satisfying these condition is empty, we provide a simple example of an instance of (3.22), which satisfies the condition of Theorem 3.4.4:

Example 3.4.1. *Let* $m = n = 1$ *and let* $\Omega_{\boldsymbol{q}} = \mathbb{T}$*. Consider the matrix valued functions* $\widetilde{\boldsymbol{\Gamma}}, \widetilde{\boldsymbol{\Sigma}}$ *defined by*

$$\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q}) = \begin{pmatrix} 0 & -(2+\cos(2\pi\boldsymbol{q})) \\ (2+\cos(2\pi\boldsymbol{q})) & 1 \end{pmatrix}, \;\; \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{q}) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Obviously, a valid choice for $\boldsymbol{Q}$ *in Proposition 3.2.2 is*

$$\boldsymbol{Q} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Moreover,

$$\boldsymbol{C} = \begin{pmatrix} 19/18 & -(1/6) \\ -(1/6) & 1 \end{pmatrix}.$$

satisfies (3.31)*. This follows by virtue of Lemma A.0.1. We provide a plot of the eigenvalues of the matrix*

$$\boldsymbol{R}(\boldsymbol{q}) = \widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q})\boldsymbol{C} + \boldsymbol{C}\widetilde{\boldsymbol{\Gamma}}^{T}(\boldsymbol{q}), \tag{3.32}$$

as a function of $\boldsymbol{q}$ *in Figure 3.1.*

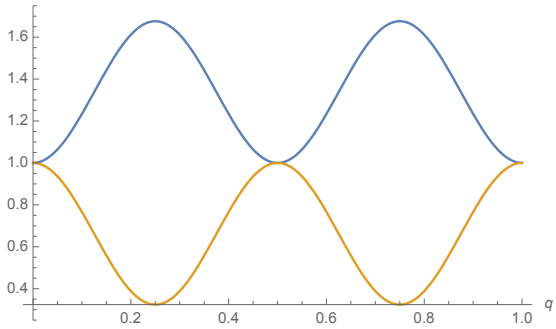

Figure 3.1: $\boldsymbol{q}$ vs. the eigenvalues of the matrix $\boldsymbol{R}(\boldsymbol{q})$ as defined in (3.32).

3.4.2 Related results

In this section we briefly review results on the ergodicity properties of hypoelliptic stochastic differential equations which are related to the results presented here. The Lyapunov based techniques on which the proofs of our ergodicity results rely have been extensively studied in the context of stochastic differential equations (see [85, 118, 81, 12]) as well as in the context of discrete time Markov chains (see e.g. [47, 84, 86, 45]). In particular, we mention the application of these techniques to prove geometric ergodicity of solutions of the underdamped Langevin equation in [118, 81, 12]. As discussed in Section 3.1.2, the structure of the SDE (3.1-3.3) resembles the structure of the underdamped Langevin equation and it is therefore not surprising that also the structure of the Lyapunov functions constructed in the proofs of [81] resemble the structure of the Lyapunov functions presented in the latter two references. Moreover, conditions for the existence of Lyapunov functions for Itô diffusion processes which are similar to certain instances of (3.1-3.3) are derived in [81]. In particular, Lemma 4.2 in that reference implies the existence of a suitable Lyapunov function under the assumptions of Theorem 3.4.2 and the additional condition that $\boldsymbol{Q} = \boldsymbol{I}_m$. We also mention [96], where the authors construct a Lyapunov for a Markovian reformulation of the GLE which in the representation (3.1-3.3) corresponds to the case where $\boldsymbol{\Gamma}_{1,1} = \boldsymbol{0}$ and $\boldsymbol{\Gamma}_{1,2}, \boldsymbol{\Gamma}_{2,1}, \boldsymbol{\Gamma}_{2,2} \in \mathbb{R}^{n\times n}$ are diagonal matrices. In the same article exponential exponential convergence of the law towards a unique invariant distribution μ in relative entropy and decay estimates of form (2.29) for $E = \left(L_0^2(\mu) \cap H^1(\mu), \|\cdot\|_{L^2(\mu)}\right)$. Ergodic properties of non-equilirium systems consisting of a chain of a finite number of oscillators whose ends are coupled to two heat baths have been studied in a series of papers [35, 34, 33, 102, 100]. In a simplified version these systems can be written as

$$
\begin{aligned}
\dot{\boldsymbol{r}}_1 &= -\gamma_1 \boldsymbol{r}_1 + \lambda_1 \boldsymbol{p}_1 + \sqrt{2\beta^{-1}\gamma_1}\dot{W}_1, \\
\dot{\boldsymbol{q}}_1 &= \boldsymbol{p}_1, \\
\dot{\boldsymbol{p}}_1 &= -\partial_{\boldsymbol{q}_1} U(\boldsymbol{q}) - \lambda_1 \boldsymbol{r}_1, \\
\dot{\boldsymbol{q}}_i &= \boldsymbol{p}_i, && i = 2, 3, \ldots, n-1, \\
\dot{\boldsymbol{p}}_i &= -\partial_{\boldsymbol{q}_i} U(\boldsymbol{q}), && i = 2, 3, \ldots, n-1, \\
\dot{\boldsymbol{q}}_n &= \boldsymbol{p}_n, \\
\dot{\boldsymbol{p}}_n &= -\partial_{\boldsymbol{q}_n} U(\boldsymbol{q}) - \lambda_2 \boldsymbol{r}_2, \\
\dot{\boldsymbol{r}}_2 &= -\gamma_2 \boldsymbol{r}_2 + \lambda_2 \boldsymbol{p}_n + \sqrt{2\beta^{-1}\gamma_2}\dot{W}_2.
\end{aligned}
\tag{3.33}
$$

where

$$
U(\boldsymbol{q}) = U_1(\boldsymbol{q}_1) + U_n(\boldsymbol{q}_n) + \sum_{i=2}^{n} \tilde{U}(\boldsymbol{q}_i - \boldsymbol{q}_{i-1}),
$$

with $U_1, U_2, \tilde{U} \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$, $\gamma_i > 0, \lambda_i > 0$ for $i = 1, 2$, and W_1, W_2 are two independent Wiener processes taking values in $\mathbb{R}$. Under certain conditions on the potential functions U_1, U_n and $\tilde{U}$, the existence of an invariant measure (stationary non-equilirium state) has been shown in [35]. Uniqueness conditions where derived in [34, 33], and exponential convergence to the invariant state was shown in [102] (see also the review paper [101] and [19]. In the latter reference slightly more general heat bath models are considered than above in (3.33)).

Exponential convergence towards a unique invariant measure is proven in [102] by showing the existence of a suitable Lyapunov function and by showing hypoellipticity

and controllability in the sense of Assumption 6. The construction of a suitable control in the proof provided therein relies on $\tilde{U}$ being strictly convex. We expect that the techniques which are used in [102] to prove the existence of a suitable Lyapunov function and the controllability of the SDE can be extended/modified to prove geometric ergodicity of GLEs which can be represented in the form (3.1-3.3). In fact it has been demonstrated in [101] that controllability in the sense of Assumption 6 of a system consisting of a chain of oscillators which are coupled to a single heat bath, can be proven by the same techniques as used in [102].

To the best of our knowledge there doesn't exist a result in the literature which would resemble Theorem 3.4.4.

3.4.3 Hypoellipticity conditions

In this section we provide criteria in the form of algebraic conditions on $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$ which ensure that (3.1-3.3) satisfies the parabolic Hörmander condition, which by Theorem 2.5.1, implies that the differential operators

$$\mathcal{L}_{\mathrm{GLE}},\ \mathcal{L}^{\dagger}_{\mathrm{GLE}},\ \partial_t - \mathcal{L}_{\mathrm{GLE}},\ \partial_t - \mathcal{L}^{\dagger}_{\mathrm{GLE}},$$

are hypoelliptic. Let in the following Proposition 3.4.1 $\Sigma_i, 1 \leq i \leq n+m$ denote the column vectors of $\mathbf{\Sigma}$, i.e.,

$$\mathbf{\Sigma} = [\Sigma_1, \dots \Sigma_{m+n}] \in \mathbb{R}^{(n+m)\times(n+m)},$$

or

$$\Sigma_i = \mathbf{\Sigma}\, \boldsymbol{e}_i,\ 1 \leq i \leq n+m,$$

where $\boldsymbol{e}_i$ denotes the i-th canonical basis vector in $\mathbb{R}^{n+m}$.

Proposition 3.4.1. *Let $\mathbf{\Gamma} \in \mathbb{R}^{(n+m)\times(n+m)}$ such that $-\mathbf{\Gamma}$ is stable and $\mathbf{\Sigma} \in \mathbb{R}^{(n+m)\times(n+m)}$. Any of the following conditions on $U \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}}, \mathbb{R})$, and $\mathbf{\Gamma}, \mathbf{\Sigma}$ are sufficient for (3.1-3.3) to satisfy the parabolic Hörmander condition.*

(i) $U(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T \boldsymbol{H}\boldsymbol{q} + \boldsymbol{h}^T q$ with $\boldsymbol{H} \in \mathbb{R}^{n\times n}, \boldsymbol{h} \in \mathbb{R}^n$. For all $\boldsymbol{q} \in \Omega_{\boldsymbol{q}}$

$$\mathbb{R}^{2n+m} = \mathrm{lin}\left(\left\{\boldsymbol{S}^k \begin{pmatrix} \boldsymbol{0} \\ \Sigma_i \end{pmatrix} :\ k \in \mathbb{N},\ 1 \leq i \leq n+m\right\}\right), \tag{3.34}$$

where

$$\boldsymbol{S} := -\begin{pmatrix} \boldsymbol{0} & -\boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{H} & \mathbf{\Gamma}_{1,1} & \mathbf{\Gamma}_{1,2} \\ \boldsymbol{0} & \mathbf{\Gamma}_{2,1} & \mathbf{\Gamma}_{2,2} \end{pmatrix} \in \mathbb{R}^{(2n+m)\times(2n+m)}.$$

(ii)

$$\mathbb{R}^{n+m} = \mathrm{lin}\left(\bigcup_{1\leq i\leq n+m} \left\{\mathbf{\Gamma}^k \Sigma_i :\ k \leq k_i\right\}\right), \tag{3.35}$$

where $k_i,\ 1 \leq i \leq n+m$ are defined as

$$k_i := \arg\max_{k\in\mathbb{N}} \boldsymbol{S}_0^k \begin{pmatrix} \boldsymbol{0} \\ \Sigma_i \end{pmatrix} \in \{\boldsymbol{0}\} \times \mathbb{R}^{n+m}, \tag{3.36}$$

with

$$\boldsymbol{S}_0 := -\begin{pmatrix} \mathbf{0} & -\boldsymbol{I}_n & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{1,1} & \boldsymbol{\Gamma}_{1,2} \\ \mathbf{0} & \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2} \end{pmatrix} \in \mathbb{R}^{(2n+m)\times(2n+m)}.$$

(iii) $\operatorname{rank}(\boldsymbol{\Sigma}_{2,2}) = m$, *and* $\operatorname{rank}(\boldsymbol{\Gamma}_{1,2}) = n$.

Proof. In the view of (2.25) the coefficients $\boldsymbol{b}_i$ are

$$\boldsymbol{b}_0(\boldsymbol{x}) = -\boldsymbol{G}\begin{pmatrix} \nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \\ \boldsymbol{z} \end{pmatrix},$$

and

$$\boldsymbol{b}_i = \beta^{-\frac{1}{2}}\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\Sigma}\boldsymbol{e}_i \end{pmatrix} \in \mathbb{R}^{2n+m}, 1 \le i \le n+m,$$

where $(\boldsymbol{e}_i)_{1\le i\le n+m}$ denotes the canonical basis in $\mathbb{R}^{n+m}$ and $\boldsymbol{G} \in \mathbb{R}^{(2n+m)\times(2n+m)}$ is as defined in (3.44). Since for $i > 0$ the coefficients $\boldsymbol{b}_i$ are constant in x, we find $[\boldsymbol{b}_i, \boldsymbol{b}_j] = \mathbf{0}$ and $[\boldsymbol{b}_0, \boldsymbol{b}_i] = -\nabla_{\boldsymbol{x}}\boldsymbol{b}_0\, \boldsymbol{b}_i$ for $i, j > 0$, where $\nabla_{\boldsymbol{x}}\boldsymbol{b}_0$ denotes the Jacobian of $\boldsymbol{b}_0$, i.e.,

$$\nabla_{\boldsymbol{x}}\boldsymbol{b}_0 = -\begin{pmatrix} \mathbf{0} & -\boldsymbol{I}_n & \mathbf{0} \\ \nabla^2_{\boldsymbol{q}} U & \boldsymbol{\Gamma}_{1,1} & \boldsymbol{\Gamma}_{1,2} \\ \mathbf{0} & \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2} \end{pmatrix},$$

and $\nabla^2_{\boldsymbol{q}} U$ denotes the Hessian of the potential function U. Therefore,

$$\mathscr{V}_1 = \{-\nabla_{\boldsymbol{x}}\boldsymbol{b}_0\, \boldsymbol{v} : \boldsymbol{v} \in \mathscr{V}_0\} \cup \mathscr{V}_0, \tag{3.37}$$

where

$$\left\{\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\Sigma}\boldsymbol{e}_i \end{pmatrix}\right\}_{i=1}^{n+m},$$

- In the case of (i) it follows that $\nabla_{\boldsymbol{x}}\boldsymbol{b}_0(\boldsymbol{x}) = \boldsymbol{S}$. In particular, since $\nabla_{\boldsymbol{x}}\boldsymbol{b}_0$ is constant in $\boldsymbol{x}$, (3.39) generalises to

  $$\mathscr{V}_{i+1} = \{\boldsymbol{S}\,\boldsymbol{v} : \boldsymbol{v} \in \mathscr{V}_i\} \cup \mathscr{V}_i,\ i \in \mathbb{N}. \tag{3.38}$$

  Since $\mathscr{V}_i$ consists only of constant functions, we have $\operatorname{lin}(\mathscr{V}_i(\boldsymbol{x})) \equiv \operatorname{lin}(\mathscr{V}_i)$ for all $\boldsymbol{x} \in \Omega_{\boldsymbol{x}}, i \in \mathbb{N}$, thus (3.39) implies the condition (3.34).

- Regarding (ii): Let $k_{\max} = \max_{1\le i\le n+m} k_i$. k_i being as defined in (3.36) together with (3.35) ensures that there is $\widetilde{\mathscr{V}} \subset \mathscr{V}_{k_{\max}}$ such that all elements in $\widetilde{\mathscr{V}}$ are constant and

  $$\operatorname{lin}\left(\widetilde{\mathscr{V}}\right) \equiv \begin{pmatrix} \mathbf{0} \\ \mathbb{R}^{n+m} \end{pmatrix}.$$

  Therefore,

  $$\mathscr{V}_{k_{\max}+1} \supset \{-\nabla\boldsymbol{b}_0\, \boldsymbol{v} : \boldsymbol{v} \in \widetilde{\mathscr{V}}\} \cup \widetilde{\mathscr{V}}, \tag{3.39}$$

  thus for all $\boldsymbol{x} \in \Omega_{\boldsymbol{x}}$

  $$\operatorname{lin}\left(\mathscr{V}_{k_{\max}+1}(\boldsymbol{x})\right) = \operatorname{lin}\left(\{-\nabla\boldsymbol{b}_0(\boldsymbol{x})\, \boldsymbol{v}(\boldsymbol{x}) : \boldsymbol{v} \in \widetilde{\mathscr{V}}\} \cup \widetilde{\mathscr{V}}(\boldsymbol{x})\right) = \mathbb{R}^{2n+m},$$

  where the latter equivalence is due to the fact that

  $$\operatorname{lin}\left(\{-\nabla\boldsymbol{b}_0(\boldsymbol{x})\boldsymbol{v}\ : \boldsymbol{v} \in B\} \cup B\right) = \mathbb{R}^{2n+m},$$

for all $\boldsymbol{x} \in \Omega_{\boldsymbol{x}}$ and any basis $B \subset \mathbb{R}^{2n+m}$ of $\{\mathbf{0}\} \times \mathbb{R}^{n+m}$.

- Regarding (iii): Since $\operatorname{lin}(\boldsymbol{\Sigma}_{2,2}) = \mathbb{R}^m$ and $\operatorname{rank}(\boldsymbol{\Gamma}_{1,2}) = n$ it follows that

$$\{\mathbf{0}\} \times \mathbb{R}^{n+m} = \operatorname{lin}\left(\left\{\begin{pmatrix}\mathbf{0}\\ \boldsymbol{\Gamma}\Sigma_i\end{pmatrix}, 1 \leq i \leq n+m\right\}\right),$$

thus the result follows by (ii).

□

3.4.4 Technical lemmas required in the proofs of ergodicity of (3.1-3.3)

In this section we provide the necessary technical lemmas to which we refer in the proofs of Theorems 3.4.1 to 3.4.3. We begin with showing the existence of a suitable class of suitable Lyapunov function in the case of a bounded configurational domain, i.e., $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n$ with $L > 0$.

Lemma 3.4.1. *Let $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n, L > 0$, $-\boldsymbol{\Gamma} \in \mathbb{R}^{(m+n)\times(n+m)}$ stable, and $U \in \mathcal{C}^\infty(L\mathbb{T}^n, \mathbb{R})$, then*

$$\mathcal{K}_l(\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s}) = \left(\boldsymbol{z}^T \boldsymbol{C} \boldsymbol{z} + U(\boldsymbol{q}) - U_{min} + 1\right)^l, \ l \in \mathbb{N},$$

where $\boldsymbol{C} \in \mathbb{R}^{n+m}$ is a symmetric positive definite matrix such that $\boldsymbol{\Gamma}^T\boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Gamma}$ is positive definite, defines a family of Lyapunov functions for the differential operator $\mathcal{L}_{GLE}$, i.e., for each $l \in \mathbb{N}$ there exist constants $a_l > 0$, $b_l \in \mathbb{R}$, such that for $\mathcal{L} = \mathcal{L}_{GLE}, \mathcal{K} = \mathcal{K}_l$, Assumption 3 holds for $a = a_l, b = b_l$.

Proof. The matrix $-\boldsymbol{\Gamma}$ being a stable ensures that there indeed exists a symmetric positive definite matrix $\boldsymbol{C}$ such that $\boldsymbol{\Gamma}^T\boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Gamma}$ is positive definite. Let

$$\lambda = \frac{\min \sigma(\boldsymbol{\Gamma}^T\boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Gamma})}{\max \sigma(\boldsymbol{C})}.$$

We first consider the case $l = 1$:

$$\begin{aligned}
(\mathcal{L}_H + \mathcal{L}_O)\mathcal{K}_1(\boldsymbol{x}) =& -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \cdot \nabla_{\boldsymbol{p}} \mathcal{K}_1(\boldsymbol{x}) + \boldsymbol{M}^{-1}\boldsymbol{p} \cdot \nabla_{\boldsymbol{q}} \mathcal{K}(\boldsymbol{x}) - \boldsymbol{\Gamma}\boldsymbol{z} \cdot \nabla_{\boldsymbol{z}}(\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z}) \\
&+ \frac{\beta^{-1}}{2} \nabla_{\boldsymbol{z}} \cdot \left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T \nabla_{\boldsymbol{z}}(\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z})\right) \\
=& -\left[2\boldsymbol{p}^T\boldsymbol{C}_{1,1} + 2\boldsymbol{s}^T\boldsymbol{C}_{1,2} - \boldsymbol{p}^T\boldsymbol{M}^{-1}\right]\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) - 2\boldsymbol{z}^T\boldsymbol{\Gamma}^T\boldsymbol{C}\boldsymbol{z} \\
&+ \beta^{-1} \sum_{i,j} \left[\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\right]_{i,j} \\
\leq& \, c_1 ||\boldsymbol{z}||_2 - 2\boldsymbol{z}^T\boldsymbol{\Gamma}^T\boldsymbol{C}\boldsymbol{z} + \beta^{-1} \sum_{i,j} \left[\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\right]_{i,j} \\
\leq& \, \frac{c_1}{\epsilon_1 \min \sigma(\boldsymbol{\Gamma}^T\boldsymbol{C})} + \epsilon_1 \boldsymbol{z}^T\boldsymbol{\Gamma}^T\boldsymbol{C}\boldsymbol{z} - 2\boldsymbol{z}^T\boldsymbol{\Gamma}^T\boldsymbol{C}\boldsymbol{z} + \beta^{-1} \sum_{i,j} \left[\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\right]_{i,j} \\
\leq& \, -(\lambda - \epsilon_1)\mathcal{K}(\boldsymbol{x}) \\
&+ \lambda(U(\boldsymbol{q}) - U_{min} + 1) + \frac{c_1}{\epsilon \min \sigma(\boldsymbol{\Gamma}^T\boldsymbol{C})} + \beta^{-1} \sum_{i,j} \left[\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\right]_{i,j} \\
=& \, -a_1\mathcal{K}_1(\boldsymbol{x}) + b_1,
\end{aligned}$$

where

$$c_1 = \max_{\boldsymbol{q}\in\Omega_{\boldsymbol{q}},\|\boldsymbol{z}\|_2=1} -[2\boldsymbol{p}^T\boldsymbol{C}_{1,1} + 2\boldsymbol{s}^T\boldsymbol{C}_{1,2} - \boldsymbol{p}^T\boldsymbol{M}^{-1}]\nabla_{\boldsymbol{q}}U(\boldsymbol{q}),$$

and

$$a_1 := (\lambda - \epsilon_1),\ b_1 := \lambda(U(\boldsymbol{q}) - U_{min} + 1) + \frac{c_1}{\epsilon_1 \min\sigma(\boldsymbol{\Gamma}^T\boldsymbol{C})} + \beta^{-1}\sum_{i,j}\left[\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\right]_{i,j},$$

with $\epsilon_1 > 0$ sufficiently small so that $a_1 > 0$.

For $l > 1$ we find:

$$\begin{aligned}
(\mathcal{L}_H + \mathcal{L}_O)\mathcal{K}_l(\boldsymbol{x}) &= l\mathcal{K}_{l-1}(\boldsymbol{x})\mathcal{L}_H\mathcal{K}_1(\boldsymbol{x}) + l\mathcal{K}_{l-1}(\boldsymbol{x})(-\boldsymbol{\Gamma}^T\boldsymbol{z}\cdot\nabla_{\boldsymbol{z}}\mathcal{K}_1(\boldsymbol{x})) \\
&\quad + l\frac{\beta^{-1}}{2}\nabla_{\boldsymbol{z}}\cdot\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\nabla_{\boldsymbol{z}}\mathcal{K}_1(\boldsymbol{x})\mathcal{K}_{l-1}(\boldsymbol{x})\right) \\
&= -l\mathcal{K}_{l-1}(\boldsymbol{x})(\boldsymbol{z}^T\boldsymbol{\Gamma}^T\boldsymbol{C}\boldsymbol{z}) + l\beta^{-1}\sum_{i,j}\left[\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\right]_{i,j}\mathcal{K}_{l-1}(\boldsymbol{x}) \\
&\quad + 2l(l-1)\beta^{-1}\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\boldsymbol{z}\mathcal{K}_{l-2}(\boldsymbol{x}) \\
&\leq -l\mathcal{K}_{l-1}(\boldsymbol{x})\left((\mathcal{L}_H + \mathcal{L}_O)\mathcal{K}_1(\boldsymbol{x}) + c_2\right) \\
&\leq l\mathcal{K}_{l-1}(\boldsymbol{x})\left(-a_1\mathcal{K}_1(\boldsymbol{x}) + b_1 + c_2\right) \\
&\leq l\left(-a_1\mathcal{K}_l(\boldsymbol{x}) + \frac{b_1 + c_2}{\epsilon_l^{l-1}} + \epsilon_l\mathcal{K}_l\right) = -a_l\mathcal{K}_l(\boldsymbol{x}) + b_l,
\end{aligned} \tag{3.40}$$

with

$$c_2 = -\beta^{-1}\sum_{i,j}\left[\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\right]_{i,j} + \beta^{-1}\sum_{i,j}\left[\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\right]_{i,j} + 2(l-1)\beta^{-1}\sup_{\boldsymbol{x}\in\mathcal{X}}\left(\frac{\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\boldsymbol{C}\boldsymbol{z}}{\mathcal{K}(\boldsymbol{x})}\right),$$

and

$$a_l := l(a_1 - \epsilon_l),\ b_l := l\frac{b_1 + c_2}{\epsilon_l^{l-1}},$$

where $\epsilon_l > 0$ sufficiently small so that $a_l > 0$. □

We next show the existence of a minorisation condition in the case of $\Omega_{\boldsymbol{q}} = \mathbb{T}^n$. The idea of the proof is to decompose the diffusion process into and Ornstein-Uhlenbeck process and a bounded remainder term, which then allows to conclude the existence of a minorising measure by virtue of the fact that the solution of Fokker-Planck equation associated with the Ornstein-Uhlenebeck process is a nondegenerated Gaussian at all times $t > 0$ and thus has full support. The idea of this approach is borrowed from [72] where it was used to show the minorisation condition for a discretised version of the underdamped Langevin equation. Other applications of this technique can be found in [99, 59].

Lemma 3.4.2. *Let $\Omega_{\boldsymbol{q}} = \mathbb{T}^n$. If $\boldsymbol{\Gamma} \in \mathbb{R}^{(n+m)\times(n+m)}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{(n+m)\times(n+m)}$ are as in Theorem 3.4.1, then Assumption 4 (minorisation condition) holds for the SDE (3.1-3.3).*

Proof of Lemma 3.4.2. Let $\boldsymbol{q}(0) = \boldsymbol{q}_0$ and $\boldsymbol{z}(0) = \boldsymbol{z}_0$ with

$$(\boldsymbol{q}_0, \boldsymbol{z}_0) \in \Omega_{\boldsymbol{q}} \times \mathcal{C}_r,$$

where

$$C_r = \{z \in \Omega_{\boldsymbol{z}} : \|z\| < r\},$$

for arbitrary but fixed $r > 0$.

We can write the solution of (3.1-3.3) as

$$\boldsymbol{z}(t) = \boldsymbol{z}_0 + \mathscr{D}_{\boldsymbol{z}}(t) + \mathscr{G}_{\boldsymbol{z}}(t), \quad \boldsymbol{q}(t) = \boldsymbol{q}_0 + \mathscr{D}_{\boldsymbol{q}}(t) + \mathscr{G}_{\boldsymbol{q}}(t), \tag{3.41}$$

with

$$\mathscr{D}_{\boldsymbol{z}}(t) = \int_0^t e^{-(t-s)\boldsymbol{\Gamma}} \begin{pmatrix} -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}(s)) \\ \boldsymbol{0} \end{pmatrix} \mathrm{d}s, \quad \mathscr{G}_{\boldsymbol{z}}(t) = \int_0^t e^{-(t-s)\boldsymbol{\Gamma}} \boldsymbol{\Sigma} \mathrm{d}\boldsymbol{W}(s),$$

and

$$\mathscr{D}_{\boldsymbol{q}}(t) = \int_0^t \Pi_{\boldsymbol{p}} \mathscr{D}_{\boldsymbol{z}}(s) \mathrm{d}s, \quad \mathscr{G}_{\boldsymbol{q}}(t) = \int_0^t \Pi_{\boldsymbol{p}} \mathscr{G}_{\boldsymbol{z}}(s) \mathrm{d}s.$$

The variables $\mathscr{G}_{\boldsymbol{q}}(t)$ and $\mathscr{G}_{\boldsymbol{z}}(t)$ are correlated and Gaussian, i.e.,

$$\begin{pmatrix} \mathscr{G}_{\boldsymbol{q}}(t) \\ \mathscr{G}_{\boldsymbol{z}}(t) \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}_t, \mathcal{V}_t),$$

with some $\boldsymbol{\mu}_t \in \Omega_{\boldsymbol{x}}$ and $\mathcal{V}_t \in \mathbb{R}^{(2n+m)\times(2n+m)}$. More specifically, $\tilde{\boldsymbol{z}}(t) = \boldsymbol{z}(0) + \mathscr{G}_{\boldsymbol{z}}(t)$ and $\boldsymbol{q}(0) + \mathscr{G}_{\boldsymbol{q}}(t)$ corresponds to the solution of of the linear SDE

$$\begin{aligned} \dot{\tilde{\boldsymbol{q}}} &= \boldsymbol{M}^{-1}\tilde{\boldsymbol{p}}, \\ \dot{\tilde{\boldsymbol{z}}} &= -\boldsymbol{\Gamma}\tilde{\boldsymbol{z}} + \boldsymbol{\Sigma}\dot{\boldsymbol{W}}, \end{aligned} \tag{3.42}$$

where $\tilde{\boldsymbol{z}}(t) = (\tilde{\boldsymbol{p}}(t), \tilde{\boldsymbol{s}}(t)) \in \Omega_{\boldsymbol{p}} \times \Omega_{\boldsymbol{s}}$. The law of $\tilde{\boldsymbol{q}}(t), \tilde{\boldsymbol{z}}(t)$ has full support for all $t > 0$, provided that the covariance matrix $\mathcal{V}_t$ is invertible. This is indeed the case since $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are required to be such that (3.1-3.3) satisfies the parabolic Hörmander condition. Therefore also (3.42) satisfy the parabolic Hörmander condition. By Theorem 2.5.1 it follows that the law of $(\tilde{\boldsymbol{q}}(t), \tilde{\boldsymbol{z}}(t))$ has a density with respect to the Lebesgue measure for any $t > 0$, which rules out the possibility of $\mathcal{V}_t$ being singular.

Let $\boldsymbol{C} \in \mathbb{R}^{(n+m)\times(n+m)}$ be symmetric positive definite such that $\boldsymbol{\Gamma}\boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Gamma}^T$ is positive definite as well, and consider the norm $\|\cdot\|_{\boldsymbol{C}}$,

$$\|\cdot\|_{\boldsymbol{C}} := \boldsymbol{z}^T \boldsymbol{C} \boldsymbol{z}, \ \boldsymbol{z} \in \mathbb{R}^{n+m}.$$

The increment $\mathscr{D}_{\boldsymbol{z}}(t)$ is uniformly bounded since

$$\|\mathscr{D}_{\boldsymbol{z}}(t)\|_{\boldsymbol{C}} \leq \|\boldsymbol{\Gamma}^{-1}\|_{\mathcal{B}(\boldsymbol{C})} \|\nabla_{\boldsymbol{q}} U\|_{L^\infty} < \infty,$$

where

$$\|\boldsymbol{\Gamma}^{-1}\|_{\mathcal{B}(\boldsymbol{C})} = \max_{v \in \mathbb{R}^{2n}} \frac{\|\boldsymbol{\Gamma}^{-1} v\|_{\boldsymbol{C}}}{\|v\|_{\boldsymbol{C}}} = \frac{1}{2} \min \sigma \left(\boldsymbol{\Gamma}^T \boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Gamma} \right),$$

denotes the operator norm of $\boldsymbol{\Gamma}^{-1}$ induced by $\|\cdot\|_{\boldsymbol{C}}$. It follows, that also $\mathscr{D}_{\boldsymbol{q}}(t)$ is bounded since

$$\|\mathscr{D}_{\boldsymbol{q}}(t)\| \leq t \|\mathscr{D}_{\boldsymbol{z}}(t)\|_{\boldsymbol{C}} < \infty.$$

Let $\mu_{x_0,t}$ denote the law of $(\boldsymbol{q}(t), \boldsymbol{z}(t))$ and $\rho_{x_0,t}$ the associated density. For fixed $t > 0$,

the terms $\mathscr{D}_{\boldsymbol{q}}(t)$ and $\mathscr{D}_{\boldsymbol{z}}(t)$ are bounded and the law of $(\boldsymbol{q}(0)+\mathscr{G}_{\boldsymbol{q}}(t), \boldsymbol{z}(0)+\mathscr{G}_{\boldsymbol{z}}(t))$ has full support, is follows that the law $\mu_{x_0,t}$ of the superposition of these two random variables has full support, thus $\rho_{x_0,t} \in \mathcal{C}(\Omega_{\boldsymbol{x}}, \mathbb{R}_+)$. Now define $\rho \in \mathcal{C}(\Omega_{\boldsymbol{x}}, \mathbb{R}_+)$ as

$$\rho(x) := \min_{\boldsymbol{x}_0 \in \mathcal{C}_r} \rho_{x_0,t}(x).$$

By construction $\rho \in \mathcal{C}(\mathcal{C}_r, \mathbb{R}_+)$ and the associated probability measure satisfies the properties of η in Assumption 4. □

We next consider the case $\Omega_{\boldsymbol{q}} = \mathbb{R}^n$. The following Lemma 3.4.3 shows the existence of a suitable class of Lyapunov functions.

Lemma 3.4.3. *Let $\Omega_{\boldsymbol{q}} = \mathbb{R}^n$, $U \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ such that Assumption 7 holds. If*

(i) $-\boldsymbol{\Gamma} \in \mathbb{R}^{(n+m)\times(n+m)}$ is a stable matrix and $\boldsymbol{\Sigma} \in \mathbb{R}^{(n+m)\times(n+m)}$ such that

$$\boldsymbol{\Gamma}_{2,2}\boldsymbol{Q} + \boldsymbol{Q}\boldsymbol{\Gamma}_{2,2}^T$$

is positive definite with $\boldsymbol{Q}$ as defined in Proposition 3.1.1,

(ii) the potential function $U \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ satisfies Assumption 7.

Furthermore, if either

(iii) $\boldsymbol{\Gamma}_{1,1}$ is positive definite,

or

(iv) U satisfies Assumption 8,

then

$$\mathcal{K}_l(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = \left(\boldsymbol{z}^T\boldsymbol{C}_{A,B}\boldsymbol{z} + \|\boldsymbol{q}\|_2^2 + 2\langle \boldsymbol{p},\boldsymbol{q}\rangle + BD(U(\boldsymbol{q}) - u_{\min}) + 1\right)^l,\ l \in \mathbb{N}, \quad (3.43)$$

where

$$\boldsymbol{C}_{A,B} = \begin{pmatrix} B\boldsymbol{I}_n & A\boldsymbol{\Gamma}_{2,1}^T \\ A\boldsymbol{\Gamma}_{2,1} & B\boldsymbol{Q}^{-1} \end{pmatrix} \in \mathcal{S}_+^{n+m},$$

for suitably chosen scalars $A, B > 0$, defines a family of Lyapunov functions for the differential operator $\mathcal{L}_{\mathrm{GLE}}$, i.e., for each $l \in \mathbb{N}$ there exist constants $a_l > 0$, $b_l \in \mathbb{R}$, such that for $\mathcal{L} = \mathcal{L}_{\mathrm{GLE}}, \mathcal{K} = \mathcal{K}_l$, Assumption 3 holds for $a = a_l, b = b_l$.

Proof. Rewriting $\mathcal{K}_l$ as

$$\mathcal{K}_l(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = \left(\boldsymbol{x}^T\hat{\boldsymbol{C}}_{A,B}\boldsymbol{x} + BD(U(\boldsymbol{q}) - u_{\min}) + 1\right)^l,\ l \in \mathbb{N},$$

where

$$\hat{\boldsymbol{C}}_{A,B} = \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{I}_n & B\boldsymbol{I}_n & A\boldsymbol{\Gamma}_{2,1}^T \\ \boldsymbol{0} & A\boldsymbol{\Gamma}_{2,1} & B\boldsymbol{Q}^{-1} \end{pmatrix} \in \mathcal{S}_+^{n+m},$$

we find by successive application of Lemma A.0.1, that for any $A' \geq 0$ there exists $B' > 0$ so that for $A = A'$ and $B \geq B'$ the matrix $\hat{\boldsymbol{C}}_{A,B}$ is positive definite and thus $\mathcal{K}_l \geq 1$ and $\mathcal{K}_l(\boldsymbol{x}) \to \infty$ as $\|\boldsymbol{x}\| \to \infty$. We first consider the case $l = 1$.

$$\boldsymbol{G} := \begin{pmatrix} \boldsymbol{0} & -\boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{I}_n & \boldsymbol{\Gamma}_{1,1} & \boldsymbol{\Gamma}_{1,2} \\ \boldsymbol{0} & \boldsymbol{\Gamma}_{2,1} & \boldsymbol{\Gamma}_{2,2} \end{pmatrix} \in \mathbb{R}^{(2n+m)\times(2n+m)}, \quad (3.44)$$

and

$$\tilde{\boldsymbol{Q}} = \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix},$$

we find

$$\begin{aligned} \mathcal{L}_{\mathrm{GLE}}\mathcal{K}_1(\boldsymbol{x}) = &- (\nabla_{\boldsymbol{q}} U(\boldsymbol{q})^T, \boldsymbol{p}^T, \boldsymbol{s}^T)\boldsymbol{G}^T\hat{\boldsymbol{C}}_{A,B}\boldsymbol{x} + DB\boldsymbol{I}_n\boldsymbol{p}\cdot\nabla_{\boldsymbol{q}}U(\boldsymbol{q}) \\ &+ \frac{\beta^{-1}}{2}\nabla_{\boldsymbol{z}}\cdot\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\nabla_{\boldsymbol{z}}(\boldsymbol{z}\tilde{Q}^{-1}\boldsymbol{z})\right), \end{aligned}$$

with

$$\boldsymbol{G}^T\boldsymbol{C} = -\begin{pmatrix} \boldsymbol{I}_n & B\boldsymbol{I}_n & \boldsymbol{\Gamma}_{2,1} \\ -\boldsymbol{I}_n + \boldsymbol{\Gamma}_{1,1} & -\boldsymbol{I}_n + B\boldsymbol{\Gamma}_{1,1} + A\boldsymbol{\Gamma}_{2,1}^T\boldsymbol{\Gamma}_{2,1} & B\boldsymbol{Q}^{-1}\boldsymbol{\Gamma}_{2,1}^T \\ \boldsymbol{\Gamma}_{1,2}^T & \boldsymbol{\Gamma}_{2,1}\boldsymbol{\Gamma}_{2,2} + B\boldsymbol{\Gamma}_{1,2}^T & \boldsymbol{\Gamma}_{2,1}\boldsymbol{\Gamma}_{1,2} + B\boldsymbol{Q}^{-1}\boldsymbol{\Gamma}_{2,2}^T \end{pmatrix}.$$

Hence, by virtue of (3.29)

$$\begin{aligned} &\mathcal{L}_{GLE}\mathcal{K}_1(\boldsymbol{x}) \leq \\ &- \boldsymbol{x}^T \underbrace{\begin{pmatrix} E\boldsymbol{I}_n & \boldsymbol{0} & \boldsymbol{0} \\ (-\boldsymbol{I}_n + \boldsymbol{\Gamma}_{1,1}) & -\boldsymbol{I}_n + B\boldsymbol{\Gamma}_{1,1} + A\boldsymbol{\Gamma}_{2,1}^T\boldsymbol{\Gamma}_{2,1} & B\boldsymbol{Q}^{-1}\boldsymbol{\Gamma}_{2,1}^T \\ \boldsymbol{\Gamma}_{1,2}^T & A\boldsymbol{\Gamma}_{2,1}\boldsymbol{\Gamma}_{2,2} + B\boldsymbol{\Gamma}_{1,2}^T & A\boldsymbol{\Gamma}_{2,1}\boldsymbol{\Gamma}_{1,2} + B\boldsymbol{Q}^{-1}\boldsymbol{\Gamma}_{2,2}^T \end{pmatrix}}_{=:\,\widehat{\boldsymbol{R}}_{A,B}} \boldsymbol{x} \\ &- A\nabla_{\boldsymbol{q}}U(\boldsymbol{q})^T\boldsymbol{\Gamma}_{2,1}^T\boldsymbol{s} + F + \frac{\beta^{-1}}{2}\sum_{i,j}[\tilde{\boldsymbol{Q}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\tilde{\boldsymbol{Q}}^{-1}]_{i,j}. \end{aligned} \tag{3.45}$$

In order to show the existence of constants a_1 and b_1 such that the respective Lyapunov inequality satisfied, one needs to show that the right hand side of the above inequality (3.45) can be bounded from above by a negative definite quadratic form.

Case $\mathrm{rank}(\boldsymbol{\Gamma}_{1,1}) = n$**:** Let $A = 0$. In this case it is sufficient to show that the symmetric part

$$\widehat{\boldsymbol{R}}_{A,B}^S = \frac{1}{2}\left(\widehat{\boldsymbol{R}}_{A,B} + \widehat{\boldsymbol{R}}_{A,B}^T\right)$$

of $\widehat{\boldsymbol{R}}_{A,B}$ is positive definite. The lower right block

$$\left[\widehat{\boldsymbol{R}}_{A,B}^S\right]_{(n+1):(2n+m),(n+1):(2n+m)} = -\boldsymbol{I}_n + \frac{B}{2}\left(\boldsymbol{\Gamma}\tilde{\boldsymbol{Q}} + \tilde{\boldsymbol{Q}}\boldsymbol{\Gamma}^T\right) \in \mathbb{R}^{(n+m)\times(n+m)},$$

of $\widehat{\boldsymbol{R}}_{0,B}^s$ is positive definite for sufficiently large $B > 0$. In particular

$$\min\sigma\left(\left[\widehat{\boldsymbol{R}}_{A,B}^S\right]_{(n+1):(2n+m),(n+1):(2n+m)}\right) = O(B),$$

as $B \to \infty$. Thus, by virtue of Lemma A.0.1 for $E > 0$ there is $B' > 0$ so that that $\widehat{\boldsymbol{R}}_{0,B}^s$ is indeed positive definite for all $B \geq B^{'}$.

Case $\boldsymbol{\Gamma}_{1,1} = \mathbf{0}$: If Assumption 8 holds, then by Remark 3.4.1 this implies that there is $\overline{H} > 0$ and $\overline{h} \in \mathbb{R}$ so that

$$|\langle \boldsymbol{g}, \nabla_{\boldsymbol{q}} U(\boldsymbol{q})\rangle| \leq \overline{H}|\langle \boldsymbol{g}, \boldsymbol{q}\rangle| + \overline{h},$$

Therefore, it is sufficient to show that there are constants A, B, E so that the function

$$\begin{aligned}\varphi(\boldsymbol{x}) &= \max\left(-\boldsymbol{x}^T \widehat{\boldsymbol{R}}_{A,B}\boldsymbol{x} - A\overline{H}\boldsymbol{q}^T\boldsymbol{\Gamma}_{2,1}^T\boldsymbol{s},\ \ -\boldsymbol{x}^T \widehat{\boldsymbol{R}}_{A,B}\boldsymbol{x} + A\overline{H}\boldsymbol{q}^T\boldsymbol{\Gamma}_{2,1}^T\boldsymbol{s}\right) \\ &= \max_{i=1,2} -\boldsymbol{x}^T \tilde{\boldsymbol{R}}^{(i)}_{A,B,E}\boldsymbol{x},\end{aligned} \tag{3.46}$$

can be bounded from above by a negative definite quadratic form. This means that we have to show that for suitable constants $A, B, E > 0$ the symmetric part of the matrix

$$\tilde{\boldsymbol{R}}^{(i)}_{A,B,E} = \begin{pmatrix} E\boldsymbol{I}_n & \mathbf{0} & (-1)^i A\overline{H}\boldsymbol{\Gamma}_{2,1}^T \\ -\boldsymbol{I}_n & -\boldsymbol{I}_n + A\boldsymbol{\Gamma}_{2,1}^T\boldsymbol{\Gamma}_{2,1} & \mathbf{0} \\ \boldsymbol{\Gamma}_{1,2}^T & A\boldsymbol{\Gamma}_{2,1}\boldsymbol{\Gamma}_{2,2} & A\boldsymbol{\Gamma}_{2,1}\boldsymbol{\Gamma}_{1,2} + B\boldsymbol{Q}^{-1}\boldsymbol{\Gamma}_{2,2}^T \end{pmatrix},$$

is positive definite for $i \in \{0, 1\}$. (Note that we used $\boldsymbol{\Gamma}_{1,2}^T - \boldsymbol{Q}^{-1}\boldsymbol{\Gamma}_{2,1} = \mathbf{0}$ in the derivation of the form of $\tilde{\boldsymbol{R}}^i_{A,B}$.) Since $\boldsymbol{\Gamma}_{2,1}^T\boldsymbol{\Gamma}_{2,1}$ is positive definite we can choose A sufficiently large so that $-\boldsymbol{I}_n + A\boldsymbol{\Gamma}_{2,1}^T\boldsymbol{\Gamma}_{2,1}$ is positive definite. The positive definiteness of the symmetric part of $\tilde{\boldsymbol{R}}^{(i)}_{A,B,E}, i \in \{0, 1\}$ follows for sufficiently large $B > 0$ and $E > 0$ by successive application of Lemma A.0.1.

For $l > 1$ we find:

$$\begin{aligned}(\mathcal{L}_H + \mathcal{L}_O)\mathcal{K}_l(\boldsymbol{x}) &= l\mathcal{K}_{l-1}(\boldsymbol{x})\mathcal{L}_H\mathcal{K}_1(\boldsymbol{x}) + l\mathcal{K}_{l-1}(\boldsymbol{x})(-\boldsymbol{\Gamma}^T\boldsymbol{z}\cdot\nabla_{\boldsymbol{z}}\mathcal{K}_1(\boldsymbol{x})) \\ &\quad + l\frac{\beta^{-1}}{2}\nabla_{\boldsymbol{z}}\cdot\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\nabla_{\boldsymbol{z}}\mathcal{K}_1(\boldsymbol{x})\mathcal{K}_{l-1}(\boldsymbol{x})\right) \\ &= -l\mathcal{K}_{l-1}(\boldsymbol{x})(\boldsymbol{z}^T\boldsymbol{\Gamma}^T\tilde{\boldsymbol{Q}}\boldsymbol{z}) + l\beta^{-1}\sum_{i,j}\left[\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\tilde{\boldsymbol{Q}}\right]_{i,j}\mathcal{K}_{l-1}(\boldsymbol{x}) \\ &\quad + 2l(l-1)\beta^{-1}\boldsymbol{z}^T\tilde{\boldsymbol{Q}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\tilde{\boldsymbol{Q}}\boldsymbol{z}\mathcal{K}_{l-2}(\boldsymbol{x}) \\ &\leq -l\mathcal{K}_{l-1}(\boldsymbol{x})\left((\mathcal{L}_H + \mathcal{L}_O)\mathcal{K}_1(\boldsymbol{x}) + c_2\right) \\ &\leq l\mathcal{K}_{l-1}(\boldsymbol{x})\left(-a_1\mathcal{K}_1(\boldsymbol{x}) + b_1 + c_2\right) \\ &\leq l\left(-a_1\mathcal{K}_l(\boldsymbol{x}) + \frac{b_1 + c_2}{\epsilon_l^{l-1}} + \epsilon_l\mathcal{K}_l\right) = -a_l\mathcal{K}_l(\boldsymbol{x}) + b_l\end{aligned} \tag{3.47}$$

with

$$c_2 = -\beta^{-1}\sum_{i,j}\left[\tilde{\boldsymbol{Q}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\tilde{\boldsymbol{Q}}\right]_{i,j} + \beta^{-1}\sum_{i,j}\left[\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\tilde{\boldsymbol{Q}}\right]_{i,j}$$

and

$$a_l := l(a_1 - \epsilon_l),\ b_l := l\frac{b_1 + c_2}{\epsilon_l^{l-1}}$$

where $\epsilon_l > 0$ sufficiently small so that $a_l > 0$. □

We mention that Assumption 7 is commonly also required for the construction of suitable Lyapunov functions in the case of the underdamped Langevin equation if $\Omega_{\boldsymbol{q}}$ is unbounded. Assumption 8 an additional constraint on the potential function U, which

is not required in the case of the underdamped Langevin equation. It is therefore not surprising that this assumption can be dropped if the noise process $\boldsymbol{\eta}$ in the GLE contains a nondegenerated white noise component.

If $\boldsymbol{\Sigma}$ has full rank the minorisation can be shown by a simple control argument.

Lemma 3.4.4. *Let $\Omega_{\boldsymbol{q}} = \mathbb{R}^n$. If $\operatorname{rank}(\boldsymbol{\Sigma}) = n + m$, then (3.1-3.3) satisfies a minorisation condition (Assumption 4).*

Proof. Note that by Proposition 3.4.1, (ii) $\operatorname{rank}(\boldsymbol{\Sigma}) = n + m$ immediately implies that the SDE satisfies the parabolic Hörmander condition. Since $\boldsymbol{\Sigma}$ is invertible, we can easily solve the associated control problem which then by corollary 2.6.1 implies that a minorisation condition is satisfied. The proof of the of the existence of a suitable control is essentially the same as in the case of the underdamped Langevin equation (see e.g. [81]): Let $T > 0$ and $(\boldsymbol{q}^-, \boldsymbol{p}^-, \boldsymbol{s}^-), (\boldsymbol{q}^+, \boldsymbol{p}^+, \boldsymbol{s}^+) \in \mathbb{R}^{2n+m}$. We need to show that there exists $u \in L^1([0,T], \mathbb{R}^m)$, solving the control problem

$$\begin{aligned} \dot{\boldsymbol{q}} &= \boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= -\nabla U(\boldsymbol{q}) - \boldsymbol{\Gamma}_{1,1}\boldsymbol{p} + \boldsymbol{\Gamma}_{1,2}\boldsymbol{s} + \boldsymbol{\Sigma}_1 \boldsymbol{u}, \\ \dot{\boldsymbol{s}} &= -\boldsymbol{\Gamma}_{2,1}\boldsymbol{p} + \boldsymbol{\Gamma}_{2,2}\boldsymbol{s} + \boldsymbol{\Sigma}_2 \boldsymbol{u}, \end{aligned} \tag{3.48}$$

subject to

$$(\boldsymbol{q}(0), \boldsymbol{p}(0), \boldsymbol{s}(0)) = (\boldsymbol{q}^-, \boldsymbol{p}^-, \boldsymbol{s}^-),\ (\boldsymbol{q}(T), \boldsymbol{p}(T), \boldsymbol{s}(T)) = (\boldsymbol{q}^+, \boldsymbol{p}^+, \boldsymbol{s}^+).$$

It is easy to verify that there exists a smooth paths $\tilde{\boldsymbol{q}} \in \mathcal{C}^2([0,T], \mathbb{R}^n)$ and $\tilde{\boldsymbol{s}} \in \mathcal{C}^2([0,T], \mathbb{R}^m)$ such that

$$(\tilde{\boldsymbol{q}}(0), \dot{\tilde{\boldsymbol{q}}}(0)) = (\boldsymbol{q}^-, \boldsymbol{p}^-),\ (\tilde{\boldsymbol{q}}(T), \dot{\tilde{\boldsymbol{q}}}(T)) = (\boldsymbol{q}^+, \boldsymbol{p}^+),$$

and

$$\tilde{\boldsymbol{s}}(0) = \boldsymbol{s}^-,\ \tilde{\boldsymbol{s}}(T) = \boldsymbol{s}^+.$$

Rewrite (3.48) as a second order differential equation in $\boldsymbol{q}$ and $\boldsymbol{s}$:

$$\begin{aligned} \ddot{\boldsymbol{q}} &= -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) - \boldsymbol{\Gamma}_{1,2}\dot{\boldsymbol{q}} - \boldsymbol{\Gamma}_{1,2}\boldsymbol{s} + \boldsymbol{\Sigma}_1 \boldsymbol{u}, \\ \dot{\boldsymbol{s}} &= -\boldsymbol{\Gamma}_{2,1}\dot{\boldsymbol{q}} - \boldsymbol{\Gamma}_{2,2}\boldsymbol{s} + \boldsymbol{\Sigma}_2 \boldsymbol{u}, \end{aligned}$$

thus,

$$\boldsymbol{u}(t) = \boldsymbol{\Sigma}^{-1} \begin{pmatrix} \ddot{\tilde{\boldsymbol{q}}}(t) + \nabla_{\boldsymbol{q}} U(\tilde{\boldsymbol{q}}(t)) + \boldsymbol{\Gamma}_{1,1}\dot{\tilde{\boldsymbol{q}}}(t) + \boldsymbol{\Gamma}_{1,2}\tilde{\boldsymbol{s}}(t) \\ \dot{\tilde{\boldsymbol{s}}}(t) + \boldsymbol{\Gamma}_{2,1}\dot{\tilde{\boldsymbol{q}}}(t) + \boldsymbol{\Gamma}_{2,2}\tilde{\boldsymbol{s}}(t) \end{pmatrix}, \tag{3.49}$$

is a solution of (3.48). □

The following Lemma 3.4.5 shows that the minorisation condition is satisfied in the case of an GLE with unbounded configurational domain and $\boldsymbol{\Gamma}_{1,1} = \mathbf{0}$.

Lemma 3.4.5. *Under the same conditions as Theorem 3.4.3 if follows that Assumption 4 is satisfied for (3.1-3.3).*

Proof. By Assumption 8 the potential function U can be decomposed as

$$U(\boldsymbol{q}) = U_1(\boldsymbol{q}) + U_2(\boldsymbol{q}),$$

where $U_1 \in \mathcal{C}^\infty(\mathbb{R}^n, \mathbb{R})$ has bounded derivatives and

$$U_2(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T \boldsymbol{H}\boldsymbol{q},$$

with $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ being a positive definite matrix. Consider the dynamics

$$\begin{aligned} \dot{\boldsymbol{q}}^a &= \boldsymbol{p}^a, \\ \dot{\boldsymbol{p}}^a &= -\boldsymbol{H}\boldsymbol{q}^a - \boldsymbol{\Gamma}_{1,2}\boldsymbol{s}^a, \\ \dot{\boldsymbol{g}}^a &= -\boldsymbol{\Gamma}_{2,1}\boldsymbol{p}^a - \boldsymbol{\Gamma}_{2,2}\boldsymbol{s}^a + \frac{\beta^{-1}}{2}\boldsymbol{\Sigma}_2\dot{\boldsymbol{W}}, \\ &\text{with } \ (\boldsymbol{q}^a(0), \boldsymbol{p}^a(0), \boldsymbol{s}^a(0)) = \boldsymbol{x}_0, \end{aligned} \tag{3.50}$$

where $\boldsymbol{x}_0 \in \mathbb{R}^{2n+m}$. The solution of (3.50) is Gaussian hence

$$\mu_t^a(\mathrm{d}\boldsymbol{x}) = \mathcal{N}(\mathrm{d}\boldsymbol{x}; \boldsymbol{\mu}_t, \mathscr{V}_t),$$

where $\boldsymbol{\mu}_t \in \mathbb{R}^{2n+m}$ and $\mathscr{V}_t \in \mathbb{R}^{(2n+m)\times(2n+m)}$. Moreover, by Proposition 3.4.1, (iii), the SDE (3.50) is hypoelliptic, hence $\mathscr{V}_t$ is non-singular for all $t > 0$. As a consequence

$$\mathrm{supp}(\mu_t^a) = \Omega_{\boldsymbol{x}}.$$

Moreover we notice that

$$-\nabla_{\boldsymbol{q}} U_1(\boldsymbol{q}) = \boldsymbol{u}(\boldsymbol{q})\boldsymbol{\Sigma}_2,$$

with

$$\boldsymbol{u}(\boldsymbol{q}) = \nabla_{\boldsymbol{q}} U_1(\boldsymbol{q})\boldsymbol{I}_{n,m}\boldsymbol{\Sigma}_2^{-1},$$

where

$$\boldsymbol{I}_{n,m} = (\boldsymbol{I}_n, \boldsymbol{0}) \in \mathbb{R}^{n\times m}.$$

Using Lemma 3.4.9 it follows by the same chain of arguments as in the proof of Lemma 3.4.8, that $\boldsymbol{u}$ satisfies Novikov's condition and by virtue of Girsanov's theorem the support of the law μ_t of the solution of (3.1-3.3)with initial condition $\boldsymbol{x}(0) = \boldsymbol{x}_0$ coincides with the law of $\mu^a_{\boldsymbol{x}_0,t}$, i.e., $\mathrm{supp}(\mu_t) = \Omega_{\boldsymbol{x}}$. Let $\mu_{\boldsymbol{x}_0,t}(\mathrm{d}\boldsymbol{x}) = \rho_{\boldsymbol{x}_0,t}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$. As in the proof of Lemma 3.4.2 we can construct a minoring measure $\eta(\mathrm{d}x) = \rho(x)\mathrm{d}x$, as

$$\rho(x) := \min_{\boldsymbol{x}_0 \in \mathcal{C}_r} \rho_{x_0,t}(x).$$

where $\mathcal{C}_r \subset \mathbb{R}^{2n+m}$ is a sufficiently large compact set. □

Lemma 3.4.9 allows to conclude that Novikov's condition is satisfied under the assumptions of the preceding Lemma 3.4.5.

Lemma 3.4.6. *Let*

$$\widehat{\mathcal{K}}_\theta(\boldsymbol{x}) = e^{\frac{\theta}{2}\mathcal{K}_l(\boldsymbol{x})}, \ l = 1,$$

with $\mathcal{K}_1$ as defined in (3.43)*. Under the same conditions as in Lemma 3.4.3, and provided that Assumption 3 holds for $\mathcal{L} = \mathcal{L}_{\mathrm{GLE}}$, $\mathcal{K} = \mathcal{K}_1$, then also $\widehat{\mathcal{K}}_\theta$ satisfies Assumption 3 for $\mathcal{L} = \mathcal{L}_{\mathrm{GLE}}$ and sufficiently small $\theta > 0$.*

Proof. A simple calculation shows

$$\mathcal{L}_{\mathrm{GLE}}\widehat{\mathcal{K}}_\theta(\boldsymbol{x}) = \left(\theta\mathcal{L}_{\mathrm{GLE}}\mathcal{K}_1(\boldsymbol{x}) + \frac{\beta^{-1}}{2}\Big(\theta\sum_{i,j}\Big[(\tilde{\boldsymbol{Q}} - \boldsymbol{I}_{n+m})\widetilde{\boldsymbol{C}}\Big]_{i,j} + \theta^2\boldsymbol{z}^T\widetilde{\boldsymbol{C}}\boldsymbol{z}\Big)\right)\widehat{\mathcal{K}}_\theta(\boldsymbol{x}),$$

with

$$\widetilde{\boldsymbol{C}} = \tilde{\boldsymbol{Q}}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\tilde{\boldsymbol{Q}}^{-1}.$$

From Lemma 3.4.3 we know $\mathcal{L}_{\mathrm{GLE}}\mathcal{K}_1(\boldsymbol{x}) = O\left(-\|\boldsymbol{x}\|^2\right)$, thus

$$\mathcal{L}_{\mathrm{GLE}}\widehat{\mathcal{K}}_\theta(\boldsymbol{x}) = \left(-O\left(\theta\|\boldsymbol{x}\|^2\right) + O\left((1+\theta)\|\boldsymbol{z}\|\right) + O\left(\theta^2\|\boldsymbol{z}\|^2\right)\right)\mathcal{K}_\theta(\boldsymbol{x}),$$

thus for sufficiently small $\theta > 0$ and suitable $b \in \mathbb{R}$,

$$\mathcal{L}_{\mathrm{GLE}}\widehat{\mathcal{K}}_\theta(\boldsymbol{x}) < -\widehat{\mathcal{K}}_\theta(\boldsymbol{x}) + b.$$

□

3.4.5 Technical lemmas required in the proofs of ergodicity of (3.22)

We first show that a minorisation condition is satisfied for the (3.22) under the assumptions of Theorem 3.4.4. For $r > 0$ let in the following $C_r := \{(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) : \|\boldsymbol{p},\boldsymbol{s}\|_2 < r\}$.

Lemma 3.4.7. *Let $\Omega_{\boldsymbol{q}} = \mathbb{T}^n$ and $\widetilde{\boldsymbol{\Gamma}}_{1,2}, \widetilde{\boldsymbol{\Gamma}}_{2,1}, \widetilde{\boldsymbol{\Gamma}}_{2,2}, \widetilde{\boldsymbol{\Sigma}}_2 \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}}, \mathrm{GL}_n(\mathbb{R}))$, such that $-\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q})$ is stable for all $\boldsymbol{q} \in \Omega_{\boldsymbol{q}}$. Let $r > 0$ and $\boldsymbol{x}_0 \in C_r$. For any $t > 0$ the law $\mu_t^{\boldsymbol{x}_0} = e^{t\mathcal{L}^\dagger}\delta_{\boldsymbol{x}_0}$ of the solution $\boldsymbol{x}(t)$ of (3.22) with initial condition $\boldsymbol{x}(t) = \boldsymbol{x}_0$ has full support. In particular, Assumption 4 (minorisation condition) holds.*

Proof. Let $\boldsymbol{x}_0 = (\boldsymbol{q}_0,\boldsymbol{p}_0,\boldsymbol{s}_0) \in C_r$ and $\tilde{\boldsymbol{x}}_0 = (\boldsymbol{q}_0,\boldsymbol{p}_0,\boldsymbol{g}_0)$ with $\boldsymbol{g}_0 = \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}_0)\boldsymbol{s}_0$. Consider the following cascade of modifications of (3.22):

$$\begin{aligned}
\dot{\boldsymbol{q}}^c &= \boldsymbol{M}^{-1}\boldsymbol{p}^c,\\
\dot{\boldsymbol{p}}^c &= -\nabla U(\boldsymbol{q}) - \boldsymbol{g}^c\\
\dot{\boldsymbol{g}}^c &= \sum_{i=1}^n \boldsymbol{p}_i^c\left(\partial_{\boldsymbol{q}_i}\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}^c)\right)\boldsymbol{g}^c - \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}^c)\widetilde{\boldsymbol{\Gamma}}_{2,1}(\boldsymbol{q}^c)\boldsymbol{M}^{-1}\boldsymbol{p}^c\\
&\quad - \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}^c)\widetilde{\boldsymbol{\Gamma}}_{2,2}(\boldsymbol{q}^c)\widetilde{\boldsymbol{\Gamma}}_{1,2}^{-1}(\boldsymbol{q}^c)\boldsymbol{g}^c + \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}^c)\widetilde{\boldsymbol{\Sigma}}_2(\boldsymbol{q}^c)\dot{\boldsymbol{W}}_t,\\
&\text{with}\quad (\boldsymbol{q}^c(0),\boldsymbol{p}^c(0),\boldsymbol{g}^c(0)) = \tilde{\boldsymbol{x}}_0,
\end{aligned} \tag{3.51}$$

and

$$\begin{aligned}
\dot{\boldsymbol{q}}^b &= \boldsymbol{M}^{-1}\boldsymbol{p}^b,\\
\dot{\boldsymbol{p}}^b &= -\nabla U(\boldsymbol{q}) - \boldsymbol{g}^b,\\
\dot{\boldsymbol{g}}^b &= \boldsymbol{p}^b - \boldsymbol{g}^b + \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Sigma}}_2(\boldsymbol{q}^b)\dot{\boldsymbol{W}}_t,\\
&\quad\text{with}\quad (\boldsymbol{q}^b(0),\boldsymbol{p}^b(0),\boldsymbol{g}^b(0)) = \tilde{\boldsymbol{x}}_0,
\end{aligned} \tag{3.52}$$

and

$$\begin{aligned}\dot{\boldsymbol{q}}^a &= \boldsymbol{M}^{-1}\boldsymbol{p}^a,\\ \dot{\boldsymbol{p}}^a &= -\nabla U(\boldsymbol{q}^a) - \boldsymbol{g}^a,\\ \dot{\boldsymbol{g}}^a &= \boldsymbol{p}^a - \boldsymbol{g}^a + \dot{\boldsymbol{W}},\\ &\text{with } \ (\boldsymbol{q}^a(0), \boldsymbol{p}^a(0), \boldsymbol{g}^a(0)) = \tilde{\boldsymbol{x}}_0.\end{aligned} \tag{3.53}$$

Let $\mu_t^a, \mu_t^b, \mu_t^c$ denote the law of the solution of (3.53), (3.52) and (3.51), respectively. We show that for any $t > 0$

(i) $\operatorname{supp}(\mu_t^a) = \Omega_{\boldsymbol{x}}$,

(ii) $\operatorname{supp}(\mu_t^b) = \operatorname{supp}(\mu_t^a)$,

(iii) $\operatorname{supp}(\mu_t^c) = \operatorname{supp}(\mu_t^b)$,

(iv) $\operatorname{supp}(\mu_t) = \operatorname{supp}(\mu_t^c)$,

which then immediately implies that $\operatorname{supp}(\mu_t) = \Omega_{\boldsymbol{x}}$ for $t > 0$ and the minorisation condition follows by the same arguments as in the proof of Lemma 3.4.2.

- Regarding (i): the system (3.53) satisfies the condition of Lemma 3.4.2, hence for sufficiently large $t' > 0$ the law of (3.53) at times $t \geq t'$ has full support.
- Regarding (ii): since $\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Sigma}}_2(\boldsymbol{q})$ is invertible, the controllability properties of (3.52) are identical to the controllability properties of (3.53), hence as a consequence of the Strook-Varadhan support theorem [116] the law of (3.52) and the law of (3.53) at time t' coincide. In particular, together with (i) $\operatorname{supp}(\mu_t^c) = \operatorname{supp}(\mu_t^b) = \Omega_{\boldsymbol{x}}$.
- Regarding (iii): We show this using Theorem 2.6.3 (Girsanov's theorem). The difference of the drift terms in (3.52) and (3.51) can be written as

  $$\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}^c)\widetilde{\boldsymbol{\Sigma}}_2(\boldsymbol{q}^c)\boldsymbol{u}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{g})$$

  with $\boldsymbol{u}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{g})$ as defined in (3.56). By Lemma 3.4.8 the function $\boldsymbol{u}$ satisfies Novikov's condition (2.44), which means that Theorem 2.6.3 (Girsanov's theorem) is applicable and it follows that the support of the solution of (3.52) at t' coincides with the support of the solution of (3.51) at t', i.e., $\operatorname{supp}(\mu_t^c) = \operatorname{supp}(\mu_t^b) = \Omega_{\boldsymbol{x}}$.
- Regarding (iv): We first note that since (i)-(iii) holds, it trivially follows that $\mu_t^c = \Omega_{\boldsymbol{x}}$. Applying the change of variables $\boldsymbol{s} = \widetilde{\boldsymbol{\Gamma}}_{1,2}^{-1}(\boldsymbol{q})\boldsymbol{g}$ to (3.51) we obtain (3.22), which means that μ_t is the push-forward of μ_t^c under the map,

  $$f : \begin{pmatrix}\boldsymbol{q}\\ \boldsymbol{p}\\ \boldsymbol{g}\end{pmatrix} \mapsto \begin{pmatrix}\boldsymbol{q}\\ \boldsymbol{p}\\ \widetilde{\boldsymbol{\Gamma}}_{1,2}^{-1}(\boldsymbol{q})\boldsymbol{g}\end{pmatrix}$$

  i.e.,

  $$\mu_t(A) = f(\mu_t^c)(A) = \mu_t^c\left(f^{-1}(A)\right), \ A \in \mathcal{B}(\Omega_{\boldsymbol{x}}).$$

  Since f is a smooth ono-to-one mapping, in particular surjective, and $\operatorname{supp}(\mu_t^c) = \Omega_{\boldsymbol{x}}$ we have

  $$\operatorname{supp}(\mu_t) = \operatorname{supp}\left(f(\mu_t^c)\right) = \Omega_{\boldsymbol{x}}.$$

□

The following Lemma 3.4.8 shows that Novikov's condition is satisfied for the function u required for the application of Girsanov's theorem in the above proof of Lemma 3.4.7.

Lemma 3.4.8. *Let $\Omega_{\boldsymbol{q}} = \mathbb{T}^n$ and $\widetilde{\boldsymbol{\Gamma}}$ and $\widetilde{\boldsymbol{\Sigma}}$ as in Lemma 3.4.7. Define*

$$\begin{aligned}\boldsymbol{u}_1(\boldsymbol{q},\boldsymbol{p},\boldsymbol{g}) &= \left(\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Sigma}}_2(\boldsymbol{q})\right)^{-1}\left(\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Gamma}}_{2,1}(\boldsymbol{q})\boldsymbol{p} - \boldsymbol{p} - \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Gamma}}_{2,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Gamma}}_{1,2}^{-1}(\boldsymbol{q})\boldsymbol{g} + \boldsymbol{g}\right)\\ &= \boldsymbol{G}(\boldsymbol{q})\begin{pmatrix}\boldsymbol{p}\\ \boldsymbol{g}\end{pmatrix}\end{aligned} \tag{3.54}$$

with

$$\boldsymbol{G}(\boldsymbol{q}) := \left(\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Sigma}}_2(\boldsymbol{q})\right)^{-1}\left(\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Gamma}}_{2,1}(\boldsymbol{q}) - \boldsymbol{I}_n \quad -\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Gamma}}_{2,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Gamma}}_{1,2}^{-1}(\boldsymbol{q}) + \boldsymbol{I}_n\right) \in \mathbb{R}^{n\times 2n},$$

and

$$\boldsymbol{u}_2(\boldsymbol{q},\boldsymbol{p},\boldsymbol{g}) = -\left(\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\widetilde{\boldsymbol{\Sigma}}_2(\boldsymbol{q})\right)^{-1}\sum_{i=1}^{n}\boldsymbol{p}_i\left(\partial_{\boldsymbol{q}_i}\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})\right)\boldsymbol{g}, \tag{3.55}$$

The function

$$\boldsymbol{u}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{g}) = \boldsymbol{u}_1(\boldsymbol{q},\boldsymbol{p},\boldsymbol{g}) + \boldsymbol{u}_2(\boldsymbol{q},\boldsymbol{p},\boldsymbol{g}) \tag{3.56}$$

satisfies Novikov's condition (2.44).

Proof of Lemma 3.4.8. Since

$$\|\boldsymbol{u}_1 + \boldsymbol{u}_2\|_2^2 \le 2\|\boldsymbol{u}_1\|_2^2 + 2\|\boldsymbol{u}_2\|_2^2,$$

it is sufficient to show that Novikov's condition holds for $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. We only show Novikov's condition explicitly for $\boldsymbol{u}_1$. [4]

Since $\widetilde{\boldsymbol{\Gamma}}_{1,2}, \widetilde{\boldsymbol{\Gamma}}_{2,1}, \widetilde{\boldsymbol{\Gamma}}_{2,2}$ and $\widetilde{\boldsymbol{\Sigma}}_2$ are smooth functions of $\boldsymbol{q}$ and since $\Omega_{\boldsymbol{q}}$ is compact the spectrum of $\boldsymbol{G}^T(\boldsymbol{q})\boldsymbol{G}(\boldsymbol{q})$ is uniformly bounded from above in $\boldsymbol{q}$, hence there is $\lambda_{\max} > 0$ such that

$$\lambda_{\max}^2(\|\boldsymbol{p}\|_2^2 + \|\boldsymbol{g}\|^2) \ge (\boldsymbol{p}^T, \boldsymbol{g}^T)\boldsymbol{G}^T(\boldsymbol{q})\boldsymbol{G}(\boldsymbol{q})\begin{pmatrix}\boldsymbol{p}\\ \boldsymbol{g}\end{pmatrix} = \|\boldsymbol{u}_1(\boldsymbol{q},\boldsymbol{p},\boldsymbol{g})\|^2, \tag{3.57}$$

and therefore

$$\mathbb{E}\left[\exp(\int_0^T \|\boldsymbol{u}_1(\boldsymbol{q}(t),\boldsymbol{p}(t),\boldsymbol{g}(t))\|\mathrm{d}t)\right] \le \mathbb{E}\left[\exp(\int_0^T \lambda_{\max}^2(\|\boldsymbol{p}(t)\|^2 + \|\boldsymbol{g}(t)\|^2)\mathrm{d}t)\right],$$

for any $T > 0$. Let $\epsilon < 2\tilde{\theta}/\lambda_{\max}^2$, with $\tilde{\theta} = \theta/\tilde{\lambda}_{\max}$ and $\theta > 0, \tilde{\lambda}_{\max}$ as defined in

[4] The respective proof for $\boldsymbol{u}_2$ is essentially the same with the only difference that in (3.57) we need to bound $\|\boldsymbol{u}_2\|_2^2$ by a term proportional to $\|\boldsymbol{p}\|_2^4 + \|\boldsymbol{g}\|_2^4$ instead of bounding u_2 by a term which is proportional to $\|\boldsymbol{p}\|_2^2 + \|\boldsymbol{g}\|_2^2$ as we do in the proof for $\boldsymbol{u}_1$. By choosing $l = 2$ in (3.58) the remaining steps of the proof are then exactly the same as for u_1.

Lemma 3.4.9. We find

$$\exp(\int_0^T \lambda_{\max}^2(\|\boldsymbol{p}(t)\|^2 + \|\boldsymbol{g}(t)\|^2)\mathrm{d}t) = \exp(\frac{1}{\epsilon}\int_0^T \epsilon\lambda_{\max}^2(\|\boldsymbol{p}(t)\|^2 + \|\boldsymbol{g}(t)\|^2)\mathrm{d}t)$$
$$\leq \frac{1}{\epsilon}\int_0^T \exp(\epsilon\lambda_{\max}^2(\|\boldsymbol{p}(t)\|^2 + \|\boldsymbol{g}(t)\|^2))\mathrm{d}t,$$

by Jensen's inequality, thus

$$\mathbb{E}\left[\exp(\int_0^T \|\boldsymbol{u}_1(\boldsymbol{q}(t),\boldsymbol{p}(t),\boldsymbol{g}(t))\|\mathrm{d}t)\right] \leq \mathbb{E}\left[\frac{1}{\epsilon}\int_0^T \exp(\epsilon\lambda_{\max}^2(\|\boldsymbol{p}(t)\|^2 + \|\boldsymbol{g}(t)\|^2))\mathrm{d}t\right]$$
$$= \frac{1}{\epsilon}\int_0^T \mathbb{E}\left[\exp(\epsilon\lambda_{\max}^2(\|\boldsymbol{p}(t)\|^2 + \|\boldsymbol{g}(t)\|^2))\right]\mathrm{d}t,$$

by Tonelli's theorem. Let for $\alpha > 0$,

$$\mathcal{K}_\alpha := \mathcal{K}_{\alpha,l},\ l = 1, \tag{3.58}$$

with $\mathcal{K}_{\alpha,l}$ as defined in (3.60). Using

$$\exp(\epsilon\lambda_{\max}^2(\|\boldsymbol{p}\|^2 + \|\boldsymbol{g}\|^2)) \leq \mathcal{K}_{\tilde{\theta}}(\boldsymbol{z}), \tag{3.59}$$

we conclude using Lemma 3.4.9, (3.61)

$$\frac{1}{\epsilon}\int_0^T \mathbb{E}\left[\exp(\epsilon\lambda_{\max}^2(\|\boldsymbol{p}(t)\|^2 + \|\boldsymbol{g}(t)\|^2))\right]\mathrm{d}t \leq \frac{1}{\epsilon}\int_0^T \mathbb{E}\left[\mathcal{K}_{\tilde{\theta}}(\boldsymbol{z}(t))\right]\mathrm{d}t$$
$$\leq \frac{1}{\epsilon}\int_0^T e^{-t}\mathcal{K}_\theta(\boldsymbol{p}_0, \widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}_0)\boldsymbol{g}_0) + b(1-e^{-t})\mathrm{d}t$$
$$< \infty.$$

with $b > 0$ as specified in Lemma 3.4.9. □

Lemma 3.4.9. *Let $\Omega_{\boldsymbol{q}} = \mathbb{T}^n$ and $\widetilde{\boldsymbol{\Gamma}}$ and $\widetilde{\boldsymbol{\Sigma}}$ as in Lemma 3.4.7 and let $\boldsymbol{C} \in \mathbb{R}^{2n\times 2n}$ with*

$$\min\sigma(\boldsymbol{C}) = 1,$$

be a symmetric positive definite matrix such that

$$\widetilde{\boldsymbol{\Gamma}}^T(\boldsymbol{q})\boldsymbol{C} + \boldsymbol{C}\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q}),$$

is positive definite for all $\boldsymbol{q} \in \Omega_{\boldsymbol{q}}$. For $\alpha > 0$ and $l \in \mathbb{N}$ define

$$\mathcal{K}_{\alpha,l}(\boldsymbol{p},\boldsymbol{s}) = e^{\frac{\alpha}{2}(\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z})^l} \tag{3.60}$$

There exists $\theta > 0$ such that Assumption 3 is satisfied with $\mathcal{K} = \mathcal{K}_{\theta,l}$ and $\mathcal{L} = \widetilde{\mathcal{L}}_{\mathrm{GLE}}$. Moreover, for $\tilde{\theta} = \theta/\tilde{\lambda}_{\max}$ with

$$\tilde{\lambda}_{\max} := \max_{\boldsymbol{q}\in\Omega_{\boldsymbol{q}}}\left\{|\lambda| \mid \lambda \in \sigma\left(\widetilde{\boldsymbol{\Gamma}}_{1,2}^{-1}(\boldsymbol{q})\right)\right\}$$

the expectation of $\mathcal{K}_{\tilde{\theta},l}$ as function of the solution $(\boldsymbol{q}^c,\boldsymbol{p}^c,\boldsymbol{g}^c)$ of (3.51) can be bounded

as

$$\mathbb{E}\left[\mathcal{K}_{\tilde{\theta},l}(\boldsymbol{p}^c,\boldsymbol{g}^c) \mid (\boldsymbol{p}^c(0),\boldsymbol{g}^c(0)) = (\boldsymbol{p}_0,\boldsymbol{g}_0)\right] \leq e^{-t}\mathcal{K}_{\theta,l}(\boldsymbol{p}_0,\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}_0)\boldsymbol{g}_0) + b(1-e^{-t}) + c(l,t), \tag{3.61}$$

where $b > 0$ as above and $c(l,t)$ is a finite nonnegative constant which depends on l and t with $c(l,t) = 0$ for $l = 1$ and all $t \geq 0$.

Proof of Lemma 3.4.9. We recall that the generator of (3.22) is of the form

$$\widetilde{\mathcal{L}}_{\mathrm{GLE}} = -\nabla_{\boldsymbol{q}}U(\boldsymbol{q})\cdot\nabla_{\boldsymbol{p}} + \boldsymbol{p}\cdot\nabla_{\boldsymbol{q}} - \widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q})\boldsymbol{z}\cdot\nabla_{\boldsymbol{z}} + \frac{1}{2}\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{q})\widetilde{\boldsymbol{\Sigma}}^T(\boldsymbol{q}) : \nabla^2_{\boldsymbol{p}},$$

We show the result only for the case $l = 1$. For $l > 1$ the result follows by induction. Let $\mathcal{K}_\theta = \mathcal{K}_{\theta,1}$. Applying the generator on $\mathcal{K}_\theta$ we obtain

$$\begin{aligned}
\mathcal{L}\mathcal{K}_\theta(\boldsymbol{p},\boldsymbol{s}) &= \left(-\theta\nabla_{\boldsymbol{q}}U(\boldsymbol{q})\cdot(\boldsymbol{C}_{1,1}\boldsymbol{p} + \boldsymbol{C}_{1,2}\boldsymbol{s})\right)\mathcal{K}_\theta(\boldsymbol{p},\boldsymbol{s}) \\
&\quad + \left(-\theta\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q})\boldsymbol{z}\cdot\boldsymbol{C}\boldsymbol{z} + \frac{1}{2}\left(\theta\mathrm{tr}\left(\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{q})\widetilde{\boldsymbol{\Sigma}}^T(\boldsymbol{q})\boldsymbol{C}\right) + \theta^2\boldsymbol{z}^T\boldsymbol{C}\widetilde{\boldsymbol{\Sigma}}(\boldsymbol{q})\widetilde{\boldsymbol{\Sigma}}^T(\boldsymbol{q})\boldsymbol{C}\boldsymbol{z}\right)\right)\mathcal{K}_\theta(\boldsymbol{p},\boldsymbol{s}) \\
&= \left(-O\left(\theta\|z\|^2\right) + O\left((1+\theta)\|z\|\right) + O\left(\theta^2\|z\|^2\right)\right)\mathcal{K}_\theta(\boldsymbol{p},\boldsymbol{s}) \\
&< -\mathcal{K}_\theta(\boldsymbol{p},\boldsymbol{s}) + b
\end{aligned}$$

for sufficiently small $\theta > 0$ and sufficiently large $b > 0$. Consequentially, for $\tilde{\theta} = \theta/\tilde{\lambda}_{\max}$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{K}_{\tilde{\theta}}(\boldsymbol{p}^c(t),\boldsymbol{g}^c(t)) \mid (\boldsymbol{p}^c(0),\boldsymbol{g}^c(0)) = (\boldsymbol{p}_0,\boldsymbol{g}_0)\right] \\
=&\,\mathbb{E}\left[\mathcal{K}_{\tilde{\theta}}(\boldsymbol{p}(t),\widetilde{\boldsymbol{\Gamma}}_{1,2}^{-1}(\boldsymbol{q}(t))\boldsymbol{s}(t)) \mid (\boldsymbol{p}(0),\boldsymbol{s}(0)) = (\boldsymbol{p}_0,\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}_0)\boldsymbol{g}_0)\right] \\
\leq&\,\mathbb{E}\left[\mathcal{K}_{\tilde{\theta}}(\tilde{\lambda}_{\max}\boldsymbol{p}(t),\tilde{\lambda}_{\max}\boldsymbol{s}(t)) \mid (\boldsymbol{p}(0),\boldsymbol{s}(0)) = (\boldsymbol{p}_0,\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}_0)\boldsymbol{g}_0)\right] \\
=&\,\mathbb{E}\left[\mathcal{K}_{\theta}(\boldsymbol{p}(t),\boldsymbol{s}(t)) \mid (\boldsymbol{p}(0),\boldsymbol{s}(0)) = (\boldsymbol{p}_0,\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}_0)\boldsymbol{g}_0)\right] \\
\leq&\, e^{-t}\mathcal{K}_\theta(\boldsymbol{p}_0,\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q}_0)\boldsymbol{g}_0) + b(1-e^{-t}).
\end{aligned}$$

□

The last Lemma 3.4.10 of this section provides conditions for the existence of suitable Lyapunov functions with polynomial growth for (3.22).

Lemma 3.4.10. *Let $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n, L > 0$, $-\boldsymbol{\Gamma} \in \mathbb{R}^{(m+n)\times(n+m)}$ stable, and $U \in \mathcal{C}^\infty(L\mathbb{T}^n,\mathbb{R})$. Moreover, assume that (3.31) holds and let $\boldsymbol{C}$ as specified therein.*

$$\mathcal{K}_l(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = \left(\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z} + U(\boldsymbol{q}) - U_{min} + 1\right)^l,\ l \in \mathbb{N},$$

defines a family of Lyapunov functions for the differential operator $\mathcal{L}_{\mathrm{GLE}}$, i.e., for each $l \in \mathbb{N}$ there exist constants $a_l > 0$, $b_l \in \mathbb{R}$, such that for $\mathcal{L} = \mathcal{L}_{\mathrm{GLE}}, \mathcal{K} = \mathcal{K}_l$, Assumption 3 holds for $a = a_l, b = b_l$.

Proof. The proof is very similar to the proof Lemma 3.4.1. The existence of a suitable matrix $\boldsymbol{C}$ as specified in (3.31) allows to extend all arguments in that proof with only some very small adaptions. For this reason we skip a detailed proof here. □

3.5 Limiting dynamics

The underdamed Langevin equation and the overdamped Langevin equation can be seen as limiting dynamics of the SDE (3.1-3.3) under certain scalings of the coefficients in the matrix $\mathbf{\Gamma}$. In the remainder of this thesis we denote by

$$\boldsymbol{x}^\lambda(t) = \boldsymbol{x}(\lambda t), \tag{3.62}$$

a time rescaled version of $\boldsymbol{x}$. (We write in the same way $\boldsymbol{q}^\lambda(t)$ for $\boldsymbol{q}(\lambda t)$ and $\boldsymbol{p}^\lambda(t)$ for $\boldsymbol{p}(\lambda t)$ etc.. Moreover, consider a scaling of the matrix $\mathbf{\Gamma}$ as

$$\mathbf{\Gamma}^\mu = \begin{pmatrix} \mu_1 \mathbf{\Gamma}_{1,1} & \mu_2 \mathbf{\Gamma}_{1,2} \\ \mu_2 \mathbf{\Gamma}_{2,1} & \mu_3 \mathbf{\Gamma}_{2,2} \end{pmatrix}. \tag{3.63}$$

and let $\mathbf{\Sigma}^\mu \in \mathbb{R}^{(n+m)\times(n+m)}$ be such that

$$\mathbf{\Sigma}^\mu \mathbf{\Sigma}^{\mu T} = \mathbf{\Gamma}^\mu \boldsymbol{Q} + \boldsymbol{Q} \mathbf{\Gamma}^{\mu T}.$$

We point out that in the non-Markovian version of (3.1-3.3) the scaling of $\mathbf{\Gamma}$ as in (3.63) corresponds to a scaling of the memory kernel as

$$\boldsymbol{K}^\mu(t) = \mu_2^2 \boldsymbol{K}(\mu_3^{-1} t).$$

Rescaling (3.1-3.3) in time according to (3.62) and substituting $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$ by $\mathbf{\Gamma}^\mu$ and $\mathbf{\Sigma}^\mu$, respectively, results in the following SDE

$$\begin{aligned} \mathrm{d}\boldsymbol{q}^\lambda &= \lambda \boldsymbol{p}^\lambda \, \mathrm{d}t, \\ \begin{pmatrix} \mathrm{d}\boldsymbol{p}^\lambda \\ \mathrm{d}\boldsymbol{s}^\lambda \end{pmatrix} &= \lambda \begin{pmatrix} -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}_t^\lambda) \\ \mathbf{0} \end{pmatrix} - \lambda \mathbf{\Gamma}^\mu \begin{pmatrix} \boldsymbol{p}^\lambda \\ \boldsymbol{s}^\lambda \end{pmatrix} dt + \sqrt{\lambda \beta^{-1}} \mathbf{\Sigma}^\mu d\boldsymbol{W}, \\ \text{with } \boldsymbol{x}^\lambda(0) &\sim \mu_0. \end{aligned} \tag{3.64}$$

where $\mathbf{\Sigma}^\mu \in \mathbb{R}^{(n+m)\times(n+m)}$ is such that

$$\mathbf{\Sigma}^\mu \mathbf{\Sigma}^{\mu T} = \mathbf{\Gamma}^\mu \boldsymbol{Q} + \boldsymbol{Q} \mathbf{\Gamma}^{\mu T}.$$

In what follows we revise results on the on the the following scalings in the asymptotic limit $\epsilon \to 0$:

$$\lambda = 1, \mu_1 = 0, \mu_2 = \epsilon^{-1}, \mu_3 = \epsilon^{-2} \qquad \text{(WN-limit)}$$
$$\lambda = \epsilon^{-1}, \mu_1 = \mu_2 = \mu_3 = \epsilon^{-1} \qquad \text{(OD-limit)}$$

We revise the respective results in a Let in the following

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{0} & -\boldsymbol{D}_a \\ \boldsymbol{D}_a & \boldsymbol{D}_b \end{pmatrix} \in \mathbb{R}^{2n\times 2n} \tag{3.65}$$

with $\boldsymbol{D}_a := \mathrm{diag}(a_1, a_2, \dots a_n)$, $\boldsymbol{D}_b := \mathrm{diag}(b_1, b_2, \dots b_n)$, where the entries a_i, b_i, $1 \leq i \leq n$ are positive scalars. For the scaling (WN-limit) as $\epsilon \to 0$ the dynamics of the rescaled process (3.64) tends to the solution of the standard white noise Langevin equation. This asymptotic property was shown in [96]:

We provide the exact statement in Proposition 3.5.1. Similarly, one can show that the dynamics in the scaling (OD-limit) tends to the solution of an overdamped Langevin

equation. To our knowledge this result has not been reported in the literature so far and we therefore provide a full proof of this result in proposition 3.5.2.

Proposition 3.5.1 (White noise limit)**.** *Let $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n, L > 0$. Consider the underdamped Langevin equation*

$$\begin{aligned} \dot{\boldsymbol{Q}} &= \boldsymbol{P} \\ \dot{\boldsymbol{P}} &= -\nabla_{\boldsymbol{Q}} U(\boldsymbol{Q}) - \boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1} \boldsymbol{P} + \sqrt{2}\beta^{-1/2} (\boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1})^{1/2} \dot{\boldsymbol{W}}, \end{aligned} \tag{3.66}$$

with $(\boldsymbol{Q}(0), \boldsymbol{P}(0)) \sim \nu_0$,

where ν_0 denotes the marginal in q, p of $\mu_0(\mathrm{d}q, \mathrm{d}p, \mathrm{d}s)$, and $(\boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1})^{1/2}$ denotes a matrix square root of $\boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1}$. Let $T > 0$ and λ as specified in (WN-limit)*. For any $t \in [0, T]$, $(\boldsymbol{q}^\lambda(t), \boldsymbol{p}^\lambda(t))$ converges weakly (in law) towards $(\boldsymbol{Q}(t), \boldsymbol{P}(t))$ as $\epsilon \to 0$.*

Proof. The proof of this result can be found in [96]. □

In the recent paper [80] the authors carefully study the overdamped limits of a Markovian reformulation of a GLE where some of the coefficients are as in (3.22) functions of the position variable $\boldsymbol{q}$. The following Proposition 3.5.2 is a direct consequence of Theorem IV.1 in [80]. We point out that the same limit has been also studied in [109] for the Markovian reformulation (3.1-3.3).

Proposition 3.5.2 (Overdamped limit)**.** *Let $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n, L > 0$. Consider the overdamped Langevin equation*

$$\dot{\boldsymbol{Q}} = -\nabla_{\boldsymbol{Q}} \boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1} U(\boldsymbol{Q}) + \sqrt{2}\beta^{-1/2} \boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1} \dot{\boldsymbol{W}}, \text{ with } \boldsymbol{Q}(0) \sim \nu_0, \tag{3.67}$$

where ν_0 denotes the marginal distribution in q of $\mu_0(\mathrm{d}q, \mathrm{d}p, \mathrm{d}s)$, and $(\boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1})^{1/2}$ denotes as above a matrix square root of $\boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1}$. Let $T > 0$ and λ as specified in (WN-limit)*. For any $t \in [0, T]$, $\boldsymbol{q}^\lambda(t)$ converges weakly (in law) towards $\boldsymbol{Q}(t)$ as $\epsilon \to 0$.*

Chapter 4

Numerical treatment of the generalised Langevin equation

4.1 Stochastic splitting methods

In this section we introduce and discuss numerical discretisation schemes for Markovian reformulation of the GLE with constant coefficients (3.1-3.3).

4.1.1 H-OU based splitting methods

In this section we describe how numerical scheme for the GLE can be constructed as stochastic splitting methods. We first consider a subdivision of the stochastic vector field which resembles the splitting previously used for the white noise Langevin equation in [70, 72]

$$\begin{bmatrix} \mathrm{d}\boldsymbol{q} \\ \mathrm{d}\boldsymbol{p} \\ \mathrm{d}\boldsymbol{s} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{p}\mathrm{d}t \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{=:\mathrm{A}} + \underbrace{\begin{bmatrix} \mathbf{0} \\ -\nabla U(\boldsymbol{q})\mathrm{d}t \\ \mathbf{0} \end{bmatrix}}_{=:\mathrm{B}} + \underbrace{\begin{bmatrix} \mathbf{0} \\ -\mathbf{\Gamma}\begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{s} \end{pmatrix} dt + \mathbf{\Sigma}\mathrm{d}\boldsymbol{W}_t \end{bmatrix}}_{=:\mathrm{O}}. \tag{4.1}$$

Here the equations corresponding to the "A" and "B" are

$$\begin{bmatrix} \mathrm{d}\boldsymbol{q} \\ \mathrm{d}\boldsymbol{p} \\ \mathrm{d}\boldsymbol{s} \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}\mathrm{d}t \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \tag{4.2}$$

and

$$\begin{bmatrix} \mathrm{d}\boldsymbol{q} \\ \mathrm{d}\boldsymbol{p} \\ \mathrm{d}\boldsymbol{s} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ -\nabla U(\boldsymbol{q})\mathrm{d}t \\ \mathbf{0} \end{bmatrix}, \tag{4.3}$$

respectively, and one can easily find the corresponding flow maps as

$$\Phi^A_{\Delta t} : (\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s}) \mapsto (\boldsymbol{q} + \Delta t \boldsymbol{p}, \boldsymbol{p}, \boldsymbol{s}), \tag{4.4}$$

$$\Phi^B_{\Delta t} : (\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s}) \mapsto (\boldsymbol{q}, \boldsymbol{p} - \Delta t \nabla_{\boldsymbol{q}} U(\boldsymbol{q}), \boldsymbol{s}). \tag{4.5}$$

Similarly, the "O" -part

$$\begin{bmatrix} \mathrm{d}\boldsymbol{q} \\ \mathrm{d}\boldsymbol{p} \\ \mathrm{d}\boldsymbol{s} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ -\boldsymbol{\Gamma} \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{s} \end{pmatrix} \mathrm{d}t + \boldsymbol{\Sigma} \mathrm{d}\boldsymbol{W}_t \end{bmatrix} \tag{4.6}$$

corresponds to a multidimensional Ornstein-Uhlenbeck process and hence can be solved exactly [39, 98], resulting in the stochastic evolution map

$$\Phi^O_{\Delta t} : (\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s}) \mapsto (\boldsymbol{q}, \boldsymbol{F}_{\Delta t} \, (\boldsymbol{p}, \boldsymbol{s})^T + \boldsymbol{S}_{\Delta t} \mathscr{R}), \tag{4.7}$$

where $\mathscr{R} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{n+m})$, $\boldsymbol{F}_{\Delta t} = e^{-\Delta t \boldsymbol{\Gamma}}$ and $\boldsymbol{S}_{\Delta t} \boldsymbol{S}_{\Delta t}{}^T = \beta^{-1}[\boldsymbol{I}_{n+m} - \boldsymbol{F}_{\Delta t} \boldsymbol{F}_{\Delta t}{}^T]$.

In principle one could construct integration schemes by combining the flow maps $\Phi^A_{\Delta t}, \Phi^B_{\Delta t}$ and $\Phi^O_{\Delta t}$ in any order. We refer to any such schemes as H-OU (Hamiltonian-Ornstein-Uhlenbeck) splitting methods. Our main focus in this chapter will be on symmetric methods requiring only one computation of the force $-\nabla U$ per iteration. In molecular dynamics simulations (as well as in Bayesian inference problems) the main computational costs typically lies in the computation of the gradient of U. In particular we examine schemes of the form

$$\hat{\Phi}^{\mathrm{XYZYX}}_{\Delta t} = \Phi^{\mathrm{X}}_{\Delta t/2} \circ \Phi^{\mathrm{Y}}_{\Delta t/2} \circ \Phi^{\mathrm{Z}}_{\Delta t} \circ \Phi^{\mathrm{Y}}_{\Delta t/2} \circ \Phi^{\mathrm{X}}_{\Delta t/2}, \tag{4.8}$$

where $\mathrm{X}, \mathrm{Y}, \mathrm{Z} \in \{\mathrm{A}, \mathrm{B}, \mathrm{O}\}, \mathrm{X} \neq \mathrm{Y} \neq \mathrm{Z}$. Due to the symmetry of the integration steps, such schemes exhibit second order convergence for ergodic averages. Splitting schemes of this structure applied to the white noise Langevin equation have been extensively studied [72, 70]. Also, the scheme $\hat{\Phi}^{\mathrm{OBABO}}_{\Delta t}$, gle-OBABO was previously proposed in [22] as a numerical integration method for the GLE. Our analysis in later sections shows that this choice is not optimal for sampling purposes.

We also mention that as in the case of the underdamped Langevin equation, it is possible to systematically construct numerical methods for the GLE of arbitrary high convergence order $s \in \mathbb{N}$, for ergodic averages following the approach by [18]. This is achieved by combining the flow maps $\Phi^A_{\Delta t}$ and $\Phi^B_{\Delta t}$ in such a way that they form a symplectic integrator of order s, $\Phi^H_{\Delta t}$, for the Hamiltonian part of the equation, and to combine this integrator with an update of the multi-dimensional Ornstein-Uhlenbeck process, i.e., $\hat{\Phi}^{\mathrm{GGLA}}_{\Delta t} = \Phi^{AB}_{\Delta t} \circ \Phi^O_{\Delta t}$. The class of integrators constructed in this way for white noise Langevin equations, goes by the name "geometric Langevin algorithm" (GLA) [18]. One case easily show that as for discretisations of the underdamped Langevin equation, the order of convergence of ergodic averages for schemes constructed in this way, in the case of the GLE, is given by the order of the symplectic integrator $\hat{\Phi}^H_{\Delta t}$. Symplectic integrators of arbitrary order can be constructed by employing the Suzuki-Yoshida composition technique [73, 44], albeit, for orders $s > 2$, requiring more than a single force computation.

4.1.2 Splitting methods based on alternative decompositions

Lastly, one could also think of integration schemes which are constructed via a subdivision of the vector field in the GLE as

$$\begin{bmatrix} \mathrm{d}\boldsymbol{q} \\ \mathrm{d}\boldsymbol{p} \\ \mathrm{d}\boldsymbol{s} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{p}\mathrm{d}t \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}}_{=:\mathrm{A}} + \underbrace{\begin{bmatrix} \boldsymbol{0} \\ \begin{pmatrix} -\nabla U(\boldsymbol{q}) \\ \boldsymbol{0} \end{pmatrix} \mathrm{d}t - \boldsymbol{\Gamma} \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{s} \end{pmatrix} \mathrm{d}t + \boldsymbol{\Sigma}\mathrm{d}\boldsymbol{W}_t \end{bmatrix}}_{=:\mathrm{S}}. \tag{4.9}$$

The symmetric splitting methods gle-ASA and gle-SAS resemble the stochastic position Verlet (SPV) and the stochastic velocity Verlet (SVV) methods, respectively, which have been previously proposed in [83] as integration schemes for the white noise Langevin equation.

$$\Phi^S_{\Delta t} : (\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s}) \mapsto \left(\boldsymbol{q}, \boldsymbol{F}_{\Delta t} \, (\boldsymbol{p}, \boldsymbol{s})^T + \boldsymbol{S}_{\Delta t} \mathscr{R} + \boldsymbol{\Gamma}^{-1} [\boldsymbol{I}_{n+m} - \boldsymbol{F}_{\Delta t}] (-\nabla U(\boldsymbol{q}), \boldsymbol{0})^T\right), \; \mathscr{R} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n+m}), \tag{4.10}$$

where $\boldsymbol{F}_{\Delta t}$ and $\boldsymbol{S}_{\Delta t}$ are defined as in (4.7).

4.2 Other numerical methods for the GLE

In [7] the authors propose a family of numerical integrators based on an extended variable formalism specifically designed for memory kernels, which take the form of a Prony series and vanishing cross-correlation terms,

$$\boldsymbol{K}_{ij}(t) = \begin{cases} 0 & \text{if } i \neq j, \\ \sum_{k=1}^{m} \frac{c_k}{\tau_k} e^{-|t|/\tau_k} & \text{if } i = 0. \end{cases}$$

which in the extended variable formalism (3.1-3.1) corresponds to a choice of $\boldsymbol{\Gamma}$ as, $\boldsymbol{\Gamma}_{1,1} = 0$

$$\begin{aligned} \boldsymbol{\Gamma}_{1,1} = 0, \boldsymbol{\Gamma}_{2,2} = \boldsymbol{I}_n \otimes \mathrm{diag}(1/\tau_1, \ldots, 1/\tau_m), \\ \boldsymbol{\Gamma}_{1,2} = \boldsymbol{I}_n \otimes (\sqrt{c_1/\tau_1}, \ldots, \sqrt{c_m/\tau_m}), \boldsymbol{\Gamma}_{2,1} = \boldsymbol{\Gamma}_{1,2}^T \end{aligned} \tag{4.11}$$

Multiple splitting schemes are proposed in this work. The method to which we refer as **BB3** is based on a splitting of the form

$$\begin{bmatrix} \mathrm{d}\boldsymbol{q} \\ \mathrm{d}\boldsymbol{p} \\ \mathrm{d}\boldsymbol{s} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{p}\mathrm{d}t \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}}_{\mathrm{A}} + \underbrace{\begin{bmatrix} \boldsymbol{0} \\ -\nabla U(\boldsymbol{q}) - \boldsymbol{\Gamma}_{1,2}\boldsymbol{s}\mathrm{d}t \\ \boldsymbol{0} \end{bmatrix}}_{\mathrm{B}} + \underbrace{\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{\Gamma}_{2,1}\boldsymbol{p} - \boldsymbol{\Gamma}_{2,2}\boldsymbol{s}\mathrm{d}t + \boldsymbol{\Sigma}_{2,2}\mathrm{d}\boldsymbol{W}_t \end{bmatrix}}_{\mathrm{O}} \tag{4.12}$$

The integration order BAOB results in an updating sequence of the form as documented in (B.5).
The method to which we refer as **BB3b** is based on a splitting of the form

$$\begin{bmatrix} \mathrm{d}\boldsymbol{q} \\ \mathrm{d}\boldsymbol{p} \\ \mathrm{d}\boldsymbol{s} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{p}\mathrm{d}t \\ \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}}_{\mathrm{A}} + \underbrace{\begin{bmatrix} \boldsymbol{0} \\ -\nabla U(\boldsymbol{q})\mathrm{d}t \\ \boldsymbol{0} \end{bmatrix}}_{\mathrm{B}} + \underbrace{\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{\Gamma}_{1,2}\boldsymbol{s}\mathrm{d}t \\ \boldsymbol{0} \end{bmatrix}}_{\mathrm{C}} + \underbrace{\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{\Gamma}_{2,1}\boldsymbol{p} - \boldsymbol{\Gamma}_{2,2}\boldsymbol{s}\mathrm{d}t + \boldsymbol{\Sigma}_{2,2}\mathrm{d}\boldsymbol{W}_t \end{bmatrix}}_{\mathrm{O}} \tag{4.13}$$

Which when integrated in the order BACOCAB results in an updating sequence as documented in (B.6). Strictly speaking, the methods (B.5) and (B.6) are not Lie-Trotter schemes because with the choice $\alpha_k = \sqrt{\frac{(1-\theta_k)^2}{h}}$ the updating step for the extended modes $s_k, \quad k = 1, \dots m$ does not correspond to an exact solution of the Ornstein-Uhlenbeck process corresponding to the O-part in the splitting. This particular choice of α_k is however favoured by the authors of [7] over an exact solution of the Ornstein-Uhlenbeck process since under this choice both the scheme (B.5) and (B.6) remains ergodic in the white noise limit i.e. in the case $\tau_k \to 0$. Indeed, one can easily verify, that in the asymptotic limit $\tau_k \to 0, \ k = 1, \dots m$, that the scheme (B.6) is equivalent to

$$\begin{aligned} \boldsymbol{p}_i^{n+1/2} &= \boldsymbol{p}_i^n - \frac{\Delta t}{2} \nabla_{\boldsymbol{q}} U(q^n) \\ \boldsymbol{q}_i^{n+1} &= \boldsymbol{q}_i^n + \frac{\Delta t}{2} \boldsymbol{p}_i^{n+1/2} \\ \tilde{\boldsymbol{p}}_i^{n+1/2} &= -\Delta t \gamma (\boldsymbol{p}_i^{n+1/2} + \boldsymbol{p}_i^{n-1/2}) + \sqrt{2\Delta t \beta^{-1} \gamma} (\mathscr{R}_i^n + \mathscr{R}_i^{n-1}) \\ \boldsymbol{p}_i^{n+1} &= \tilde{\boldsymbol{p}}_i^{n+1/2} - \frac{\Delta t}{2} \nabla_{\boldsymbol{q}} U(q^{n+1}) \end{aligned} \tag{4.14}$$

where $\gamma = \sum_{k=1}^m c_k$ and $\mathscr{R}^n, n \in \mathbb{N}$ are i.i.d. normal distributed random variables.

In [115] the authors provide a numerical integrator for the GLE, where the memory kernel is assumed to be of the form

$$\boldsymbol{K}(t) = \sum_{k=1}^{N} c_k \cos(\alpha_k t) e^{-|t|/\tau_k},$$

where $c_k, \alpha_k, \tau_k \in \mathbb{R}$. For the sake of clarity we just consider the case of a single particle and and a memory kernel comprised of a single oscillatory mode $\boldsymbol{K}(t) = c_k \cos(\alpha_k t) e^{-|t|/\tau_k}$. The splitting on which the integrator is based reads

$$\begin{bmatrix} \dot{\boldsymbol{q}} \\ \dot{\boldsymbol{p}} \\ \dot{\boldsymbol{s}}_1 \\ \dot{\boldsymbol{s}}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{p} dt \\ 0 \\ \sqrt{c}\boldsymbol{p} - \alpha \boldsymbol{s}_2 \\ 0 \end{bmatrix}}_{\text{A}} + \underbrace{\begin{bmatrix} 0 \\ -U'(\boldsymbol{q}) dt - \sqrt{c}\boldsymbol{s}_1 \\ 0 \\ \alpha \boldsymbol{s}_1 \end{bmatrix}}_{\text{B}} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ -\frac{1}{\tau}\boldsymbol{s}_1 + \sqrt{2\beta^{-1}/\tau} \mathrm{d}W_t \\ -\frac{1}{\tau}\boldsymbol{s}_2 + \sqrt{2\beta^{-1}/\tau} \mathrm{d}W_t \end{bmatrix}}_{\text{O}} \tag{4.15}$$

The authors of [115] consider the integration sequence OBABO.

4.3 Metropolisation of GLE schemes

The GLE schemes which we introduced in the previous section can be used as proposals in a generalised Metropolis-Hastings algorithm. A scheme obtained in such a way is asymptotically exact, i.e., the invariant measure of the generated Markov chain corresponds exactly to the target measure.

Let P denote the transition kernel of the modified scheme obtained by adding an accept-reject step to the original scheme. The most natural way of metropolising a GLE scheme is by designing the Metropolis acceptance-rejection step such that a generalised detailed balance condition (see Section 1.2.5) of the form

$$\rho_{\boldsymbol{Q},\beta}(\boldsymbol{x}_n) P(\boldsymbol{x}_{n+1} | \boldsymbol{x}_n) = \rho_{\boldsymbol{Q},\beta}(\boldsymbol{x}_{n+1}) P(\boldsymbol{q}_n, -\boldsymbol{z}_n | \boldsymbol{q}_{n+1}, -\boldsymbol{z}_{n+1}),$$

is satisfied. For example, in the case of the gle-BAOAB scheme, this can be achieved by modifying one step in the algorithm as follows:

1. $(\boldsymbol{x}^*) \leftarrow \hat{\Phi}^{\mathrm{BAOAB}}_{\Delta t}(\boldsymbol{x}_n)$,
2. draw $u \sim U([0,1])$,
3. $$p_{\mathrm{acc}} \leftarrow \min\left\{1, \frac{\rho_{\boldsymbol{Q},\beta}(\boldsymbol{x}^*)}{\rho_{\boldsymbol{Q},\beta}(\boldsymbol{x}_n)} \frac{P^{BAOAB}_{\Delta t}\big(\ (\boldsymbol{q}_n, -\boldsymbol{z}_n) \mid (q^*, -\boldsymbol{z}^*)\ \big)}{P^{BAOAB}_{\Delta t}\big(\ (\boldsymbol{q}^*, \boldsymbol{z}^*) \mid (\boldsymbol{q}_n, \boldsymbol{z}_n)\ \big)}\right\},$$
4. if $u < p_{\mathrm{acc}}$, then
$$(\boldsymbol{x}_{n+1}) \leftarrow (\boldsymbol{x}^*),$$
otherwise,
$$(\boldsymbol{q}_{n+1}, \boldsymbol{z}_{n+1}) \leftarrow (\boldsymbol{q}_n, -\boldsymbol{z}_n);$$

where $P^{BAOAB}_{\Delta t}$ denotes the transition kernel associated with the gle-BAOAB spitting scheme, hence

$$\frac{P^{BAOAB}_{\Delta t}\big(\ (\boldsymbol{q}_n, -\boldsymbol{z}_n) \mid (\boldsymbol{q}^*, -\boldsymbol{z}^*)\ \big)}{P^{BAOAB}_{\Delta t}\big(\ (\boldsymbol{q}^*, \boldsymbol{z}^*) \mid (\boldsymbol{q}_n, \boldsymbol{z}_n)\big)} = \frac{g\left(\boldsymbol{F}_{\Delta t}(\ \hat{\boldsymbol{p}}_{n+1/2}, \boldsymbol{s}^*)^\top - (\boldsymbol{p}_n, \boldsymbol{s}_n)^\top\right)}{g\left(\boldsymbol{S}_{\Delta t}\mathscr{R}^k\right)},$$

with

$$g(x) = \exp(-\frac{\beta}{2} x^\top (\boldsymbol{S}_{\Delta t}{\boldsymbol{S}_{\Delta t}}^T)^{-1} x).$$

and $\boldsymbol{F}_{\Delta t}$ and $\boldsymbol{S}_{\Delta t}$ as previously defined in Section 4.1. Although we did not systematically evaluate the performance of this metropolised GLE integrator numerically, we don't expect it to be particularly useful in practice. This is for the following reasons: The dynamical properties of the continuous dynamics and its' (not metropolised) discretisation are at least for large step sizes lost in the metropolised version, since each time a proposal is rejected the sign of both the momentum variable and the auxiliary variables is reversed. This makes the scheme unsuitable in situations where the GLE is used as a dynamical model and for the same reason we don't expect the metropolised scheme to inherit the enhanced sampling properties of the continuous dynamics. Moreover, the dimensionality of the phase space is drastically increased due to the additional auxiliary variable. This is likely to result in a reduced acceptance rate at fixed step size in comparison to metroplised versions of the underdamped Langevin equation.

4.4 Ergodic properties of H-OU splitting methods

In this section we show the ergodicity of the Markov chain obtained via an H-OU splitting scheme of the markovian reformulation (3.1-3.3) of the GLE for compact position space $\Omega_{\boldsymbol{q}}$. Let in the remainder of this section $P_{\Delta t}$ denote the evolution operator associated with an H-OU splitting scheme using a stepsize of length Δt and assume that $\Omega_{\boldsymbol{q}}$ is compact, e.g., $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n$ and $U \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}}, \mathbb{R})$. Under these assumptions we formulate the following proposition:

Proposition 4.4.1. *Fix $l \in \mathbb{N}, l > 0$ and assume that $-\boldsymbol{\Gamma}$ is stable. Let $\mathcal{K}_l \in \mathcal{C}^\infty(\mathbb{T}^n, [0,\infty))$ be as defined in Lemma 3.4.1. There exists $\Delta t^* > 0$, such that the Markov chain associated with $P_{\Delta t}$ has a unique[1] invariant probability measure $\mu_{\Delta t}$,*

[1]The invariant measure generally depends on $\boldsymbol{\Gamma}$ though. See Section 4.6 for more details.

which admits a density with respect to the Lebesgue measure on $\Omega_{\boldsymbol{x}}$ and has finite moments, i.e., there exists $R > 0$ such that

$$\int_{\Omega_{\boldsymbol{x}}} \mathcal{K}_l \mathrm{d}\mu_{\Delta t} \leq R < \infty, \tag{4.16}$$

uniformly in time step Δt. There exist $\lambda, C > 0$ such that for all $\varphi \in L^\infty_{\mathcal{K}_l}$ the corresponding Markov chain converges geometrically with rate $r = e^{-\lambda \Delta t}$, i.e.,

$$\left| (P^k_{\Delta t}\varphi)(\boldsymbol{x}) - \mathbb{E}_{\mu_{\Delta t}}\varphi \right| \leq C\mathcal{K}_l(\boldsymbol{x}) r^n \|\varphi\|_{L^\infty_{\mathcal{K}_l}}, \tag{4.17}$$

for all $k \in \mathbb{N}$ and almost all $\boldsymbol{x} \in \Omega_{\boldsymbol{x}}$.

Proof. In order to prove this theorem we follow the same strategy as in [72], where geometric ergodicity is shown for a discretisation of the underdamped Langevin equation. We first show that under the condition of the theorem a uniform Lyapunov condition (Assumption 9) and a uniform minorisation condition (Assumption 10) hold for $P_{\Delta t}$. The two main statements of the theorem, namely (i) the exponential convergence to the unique invariant measure $\mu_{\Delta t}$, with a convergence rate uniform in Δt, and (ii) the finiteness of moments of $\mu_{\Delta t}$ of arbitrary high order, then follow from the application of Theorem 2.6.2 to $P^{\lceil T/\Delta t \rceil}_{\Delta t}$, with $T > 0$ sufficiently large.

Assumption 9 (Uniform Lyapunov condition)**.** *For any $l \in \mathbb{N},\ l > 0$, there exists $\Delta t^* > 0$ and $a_l, b_l > 0$ such that for any $\Delta t,\ 0 < \Delta t \leq \Delta t^*$,*

$$P_{\Delta t}\mathcal{K}_l \leq e^{-a_l \Delta t}\mathcal{K}_l + b_l \Delta t.$$

Assumption 10 (Uniform minorization condition)**.** *Consider $T > 0$ sufficiently large, and fix any $z_{\max} > 0$. There exist $\Delta t^*, \alpha > 0$ and a probability measure ν such that, for any bounded, measurable non-negative function φ, and any $0 < \Delta t < \Delta t^*$,*

$$\inf_{|\boldsymbol{z}| \leq z_{\max}} \left(P^{\lceil T/\Delta t \rceil}_{\Delta t}\varphi \right)(\boldsymbol{x}) \geq \alpha \int_{\Omega_{\boldsymbol{x}}} \varphi(x)\nu(dx).$$

We show that Assumption 9 and Assumption 10 hold for the integration sequence BAO. The respective proof is substantially shorter than a proof of Assumption 9 and Assumption 10 would be for a symmetric splitting scheme. However, in the case of a symmetric scheme the proof of Assumption 9 and Assumption 10 is conceptually not different from the proof presented below and relies on the same techniques and estimates (see Remark 4.4.1).

Uniform Lyapunov condition Let $\boldsymbol{C}$ be a symmetric positive definite matrix in $\mathbb{R}^{(n+m)\times(n+m)}$ such that $\boldsymbol{\Gamma C} + C\boldsymbol{\Gamma}^T$ is positive definite. Without loss of generality assume that

$$\min_{\boldsymbol{q} \in \mathbb{T}^n} U(\boldsymbol{q}) \geq 1.$$

Under this assumption it is sufficient to show that for any $l \in \mathbb{N}$ there exists $a_l, b_l > 0$ such that Assumption 9 is satisfied for

$$\mathcal{K}_l(\boldsymbol{x}) = (\boldsymbol{z}^T \boldsymbol{C} \boldsymbol{z})^l,$$

with $\boldsymbol{C} \in \mathbb{R}^{(n+m)\times(n+m)}$ as defined in the proof of Lemma 3.4.1. For the integration sequence BAO, the action of the corresponding stochastic flow map projected on the

$\boldsymbol{z}$-coordinates is found to be

$$\Pi_{\boldsymbol{z}} \circ \Phi_{\Delta t}^{\mathrm{BAO}}(\boldsymbol{x}) = \boldsymbol{F}_{\Delta t} \begin{pmatrix} \boldsymbol{p} - \Delta t \nabla_{\boldsymbol{q}} U(\boldsymbol{q}) \\ \boldsymbol{s} \end{pmatrix} + \boldsymbol{S}_{\Delta t} \mathscr{R} \quad \text{with} \quad \mathscr{R} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n+m}),$$

and hence the action of the evolution operator applied to $\mathcal{K}_l$ is

$$\begin{aligned} P_{\Delta t}^{\mathrm{BAO}} \mathcal{K}_l(\boldsymbol{x}) &= \mathbb{E}\left[\left(\Pi_{\boldsymbol{z}} \circ \Phi_{\Delta t}^{\mathrm{BAO}}(\boldsymbol{x}) \cdot \boldsymbol{C} \ \left(\Pi_{\boldsymbol{z}} \circ \Phi_{\Delta t}^{\mathrm{BAO}}(\boldsymbol{x}) \right) \right)^l \right], \\ &= (\boldsymbol{z}^T \boldsymbol{F}_{\Delta t}{}^T \boldsymbol{C} \boldsymbol{F}_{\Delta t} \boldsymbol{z})^l + \mathbb{E}\left[r_{\Delta t, l}(\boldsymbol{x}, \mathscr{R}) \right], \end{aligned} \tag{4.18}$$

where the remainder $r_{\Delta t,l}(\boldsymbol{x}, \mathscr{R})$ is a polynomial in $\boldsymbol{z}$ and $\mathscr{R}$ of degree $2l-1$. Furthermore, since the expectation of odd powers of $\mathscr{R}$ vanishes, one finds $\mathbb{E}\left[r_{\Delta t,l}(\boldsymbol{x}, \mathscr{R}) \right] \in o(\Delta t)$ as $\Delta t \to 0$ by virtue of Lemma 4.4.2 (vi). Therefore,

$$\tilde{r}_l(\boldsymbol{z}, \boldsymbol{q}, \Delta t) := \Delta t^{-1} \mathbb{E}\left[r_{\Delta t, l}(\boldsymbol{x}, \mathscr{R}) \right].$$

is well defined for $\Delta t = 0$ so that $\tilde{r}_l(\boldsymbol{z}, \cdot, \cdot)$ is continuos, thus bounded on the compact set $\Omega_{\boldsymbol{q}} \times [0, \Delta t^*]$ for any $\boldsymbol{z} \in \Omega_{\boldsymbol{p}} \times \Omega_{\boldsymbol{s}}$ and $\Delta t^* > 0$. By construction $\tilde{r}_l(\cdot, \boldsymbol{q}, \Delta t)$ is a polynomial of degree $2l-1$ and the smoothness and boundedness properties of $\tilde{r}_l(\boldsymbol{z}, \cdot, \cdot)$ are inherited by the $\boldsymbol{q}$ and Δt^* dependent coefficients, hence using Lemma 4.4.1 one finds that there exists $K_l, C_{l,U,\Delta t^*} > 0$ such that

$$\tilde{r}_l(\boldsymbol{z}, \boldsymbol{q}, \Delta t) \leq K_l \epsilon (\boldsymbol{z}^T \boldsymbol{C} \boldsymbol{z})^l + \frac{1}{\epsilon^{2l-1}} C_{l,U,\Delta t^*} \ \forall \epsilon > 0.$$

Using this and Lemma 4.4.2 (i) one finds

$$P_{\Delta t}^{\mathrm{BAO}} \mathcal{K}_l(\boldsymbol{x}) \leq \left(\exp(-\Delta t l \lambda) + \Delta t K_l \epsilon \right) (\boldsymbol{z}^T \boldsymbol{C} \boldsymbol{z})^l + \frac{\Delta t}{\epsilon^{2l-1}} C_{l,U,\Delta t^*},$$

with $\lambda = \frac{\sigma_{\min}[\boldsymbol{\Gamma}^T \boldsymbol{C} + \boldsymbol{C} \boldsymbol{\Gamma}]}{\sigma_{\max}(\boldsymbol{C})}$. Choosing ϵ for instance as $K_l \epsilon = 2l\lambda$ leads to

$$\exp(-\Delta t l \lambda) + \Delta t K_l \epsilon \leq \exp(-\Delta t \lambda l / 2).$$

for sufficiently small Δt. We conclude that

$$P_{\Delta t}^{\mathrm{BAO}} \mathcal{K}_l \leq e^{-a_l \Delta t} \mathcal{K}_l + b_l \Delta t,$$

with $a_l = \lambda l / 2$ and $b_l = \frac{1}{\epsilon^{2l-1}} C_{l,U,\Delta t^*}$.

Uniform minorization condition Let $\boldsymbol{x}^{(n+1)} = \Phi_{\Delta t}^{\mathrm{BAO}}(\boldsymbol{x}^{(n)})$ and $\boldsymbol{x}^{(0)}$ such that $\|\boldsymbol{z}^{(0)}\| \leq z_{\max}$. We prove that Assumption 10 holds for indicator functions of cylinder sets which is equivalent to showing

$$\mathbb{P}\left((\boldsymbol{q}^{\lceil T/\Delta t \rceil}, \boldsymbol{z}^{\lceil T/\Delta t \rceil}) \in A_{\boldsymbol{q}} \times A_{\boldsymbol{z}} \mid \|\boldsymbol{z}^{(0)}\| \leq z_{\max} \right) \leq \alpha \nu(A). \tag{4.19}$$

where $A = A_{\boldsymbol{q}} \times A_{\boldsymbol{z}}$ and $A_{\boldsymbol{q}} \in \mathscr{B}(\Omega_{\boldsymbol{q}})$ and $A_{\boldsymbol{z}} \in \mathscr{B}(\Omega_{\boldsymbol{z}})$. For general $A \in \mathscr{B}(\Omega_{\boldsymbol{x}}) = \mathscr{B}(\Omega_{\boldsymbol{q}}) \otimes \mathscr{B}(\Omega_{\boldsymbol{z}})$ the statement follows then using the same standard techniques as used in the construction of the product measure of finitely many measurable spaces (see e.g. [10]).

We follow the same strategy as in [72] to the show the validity of (4.19). That is, we rewrite the n-th iterate $(\boldsymbol{q}^n, \boldsymbol{z}^n)$ as a perturbation of the BAO solution $(\mathscr{G}_{\boldsymbol{z}}{}^n, \mathscr{G}_{\boldsymbol{q}}{}^n)$ of

the system (3.1-3.3) with $U = 0$ and initial value $(\boldsymbol{q}(0), \boldsymbol{z}(0)) = (\boldsymbol{0}, \boldsymbol{0})$, i.e.

$$\boldsymbol{z}^n = \mathscr{D}_{\boldsymbol{z}}{}^n + \mathscr{G}_{\boldsymbol{z}}{}^n, \quad \boldsymbol{q}^n = \mathscr{D}_{\boldsymbol{q}}{}^n + \mathscr{G}_{\boldsymbol{q}}{}^n. \tag{4.20}$$

and show that for sufficiently small $\Delta t^* > 0$ the remainder $\mathscr{D}_{\boldsymbol{z}}{}^n$ and $\mathscr{D}_{\boldsymbol{q}}{}^n$ and the covariance of $(\mathscr{G}_{\boldsymbol{z}}{}^n, \mathscr{G}_{\boldsymbol{q}}{}^n)$ can be uniformly controlled for $0 < \Delta t \leq \Delta t^*$. For this purpose we rewrite $(\boldsymbol{q}^n, \boldsymbol{p}^n)$ explicitly in terms of $(\boldsymbol{q}^0, \boldsymbol{p}^0)$ and the random noise terms $(\mathscr{R}^k)_{0 \leq k \leq n}$:

$$\begin{aligned}
\boldsymbol{q}^n &= \boldsymbol{q}^0 + \Delta t \boldsymbol{M}^{-1} \sum_{k=0}^{n-1} \boldsymbol{p}^k - \Delta t^2 \boldsymbol{M}^{-1} \sum_{k=0}^{n-1} \nabla_{\boldsymbol{q}} U(\boldsymbol{q}^k), \\
\boldsymbol{z}^n &= \boldsymbol{F}_{\Delta t}{}^n \boldsymbol{z}^0 - \Delta t \sum_{k=1}^{n-1} \boldsymbol{F}_{\Delta t}{}^k \nabla_{\boldsymbol{q}} U(\boldsymbol{q}^{n-k}) + \beta^{-1/2} \boldsymbol{S}_{\Delta t} \sum_{k=0}^{n-1} \boldsymbol{F}_{\Delta t}{}^{n-1-k} \mathscr{R}^k.
\end{aligned}$$

This in turn allows us to find explicit expressions for $\mathscr{G}_{\boldsymbol{q}}{}^n, \mathscr{G}_{\boldsymbol{z}}{}^n$:

$$\mathscr{G}_{\boldsymbol{q}}{}^n = \Delta t \boldsymbol{M}^{-1} \Pi_{\boldsymbol{p}} \sum_{k=0}^{n-1} \mathscr{G}_{\boldsymbol{z}}{}^k, \qquad \mathscr{G}_{\boldsymbol{z}}{}^n = \beta^{-1/2} \boldsymbol{S}_{\Delta t} \sum_{k=0}^{n-1} \boldsymbol{F}_{\Delta t}{}^{n-1-k} \mathscr{R}^k,$$

and explicit expressions for $\mathscr{D}_{\boldsymbol{q}}{}^n, \mathscr{D}_{\boldsymbol{z}}{}^n$:

$$\begin{aligned}
\mathscr{D}_{\boldsymbol{z}}{}^n &= \boldsymbol{F}_{\Delta t}{}^n \boldsymbol{z}^0 + \Delta t \mathcal{F}^n, \\
\mathscr{D}_{\boldsymbol{q}}{}^n &= \boldsymbol{q}^0 + \Delta t \boldsymbol{M}^{-1} \Pi_{\boldsymbol{p}} \left(\Delta t \sum_{k=1}^{n-1} \mathcal{F}^k + (\boldsymbol{I} - \boldsymbol{F}_{\Delta t}{}^n)(\boldsymbol{I} - \boldsymbol{F}_{\Delta t})^{-1} \boldsymbol{z}^0 \right) \\
&\quad - \Delta t^2 \boldsymbol{M}^{-1} \sum_{k=0}^{n-1} \nabla_{\boldsymbol{q}} U(\boldsymbol{q}^k),
\end{aligned}$$

where

$$\mathcal{F}^n = -\sum_{k=0}^{n-1} \boldsymbol{F}_{\Delta t}{}^{n-k} \nabla \boldsymbol{U}^k, \quad \nabla \boldsymbol{U}^n := (\nabla_{\boldsymbol{q}} U^T(\boldsymbol{q}^n), \boldsymbol{0}^T)^T.$$

We note that $\mathscr{D}_{\boldsymbol{z}}{}^n$ and $\mathscr{D}_{\boldsymbol{q}}{}^n$ depend only on the initial conditions $\boldsymbol{q}^0, \boldsymbol{z}^0$, whereas the iterates $\mathscr{G}_{\boldsymbol{z}}{}^n$ and $\mathscr{G}_{\boldsymbol{q}}{}^n$ depend solely on the noise terms added in the "O"-integration step. We now show that:

1. The perturbation terms $\mathscr{D}_{\boldsymbol{z}}$ and $\mathscr{D}_{\boldsymbol{q}}$ are uniformly bounded in Δt. That is, for any $z_{\max} > 0$ and $T > 0$, there exist $\Delta t^* > 0$ and $R > 0$ s.t.

$$\sup_{\Delta t \leq \Delta t^*} \|\mathscr{D}_{\boldsymbol{z}}{}^{\lceil T/\Delta t \rceil}\| \leq R, \quad \text{and} \quad \sup_{\Delta t \leq \Delta t^*} \|\mathscr{D}_{\boldsymbol{q}}{}^{\lceil T/\Delta t \rceil}\| \leq R.$$

2. The law of $(\mathscr{G}_{\boldsymbol{z}}{}^n, \mathscr{G}_{\boldsymbol{q}}{}^n)$ is Gaussian and for sufficiently small $\Delta t^* > 0$ the covariance

$$\mathscr{V}^{\lceil T/\Delta t \rceil} = \mathbb{E}\left[\left(\mathscr{G}_{\boldsymbol{q}}{}^{\lceil T/\Delta t \rceil}, \mathscr{G}_{\boldsymbol{z}}{}^{\lceil T/\Delta t \rceil} \right) \left(\mathscr{G}_{\boldsymbol{q}}{}^{\lceil T/\Delta t \rceil}, \mathscr{G}_{\boldsymbol{z}}{}^{\lceil T/\Delta t \rceil} \right)^T \right]$$

is non-degenerate and there exist positive definite matrices $\underline{\mathscr{V}}, \bar{\mathscr{V}}$ so that

$$\underline{\mathscr{V}} \leq \mathscr{V}^{\lceil T/\Delta t \rceil} \leq \bar{\mathscr{V}},^{2} \tag{4.21}$$

holds for $0 < \Delta t \leq \Delta t^*$.

Regarding 1, since $-\boldsymbol{\Gamma}$ is stable one has

$$\|\mathcal{F}^n\| \leq \|\sum_{k=0}^{n-1} \boldsymbol{F}_{\Delta t}{}^{n-k}\| \|\nabla_{\boldsymbol{q}} U(\boldsymbol{q}^k)\| \leq \|(\boldsymbol{I} - \boldsymbol{F}_{\Delta t})^{-1}\| \|\nabla_{\boldsymbol{q}} U\| \leq \Delta t^{-1} c_{\Delta t^*}^{-1} \|\nabla_{\boldsymbol{q}} U\|$$

for $\Delta t \leq \Delta t^*$ with Δt^* sufficiently small and $c_{\Delta t^*} \leq \lambda_{\max}(-\boldsymbol{\Gamma})$. Therefore,

$$\|\mathscr{D}_{\boldsymbol{z}}{}^{\lceil T/\Delta t \rceil}\| \leq \|\boldsymbol{F}_{\Delta t}{}^{\lceil T/\Delta t \rceil}\| \|\boldsymbol{z}^0\| + \Delta t \|\mathcal{F}^n\| \leq z_0 + c_{\Delta t^*}^{-1} \|\nabla_{\boldsymbol{q}} U\|$$

The configurational component $\mathscr{D}_{\boldsymbol{q}}{}^{\lceil T/\Delta t \rceil}$ is trivially bounded, since $\Omega_{\boldsymbol{q}}$ is assumed to be compact.
Regarding 2, since the BAO splitting method defines a weakly consistent discretisation scheme, the covariance $\mathscr{V}^{\lceil T/\Delta t \rceil}$ at time $T > 0$ of the discretised dynamics converges to the covariance matrix

$$\mathscr{V}_T = \mathbb{E}[\boldsymbol{x}(T)\boldsymbol{x}(T)^T |\ \boldsymbol{x}(0) = \boldsymbol{0}]$$

of the exact solution $\boldsymbol{x}(T)$ of the SDE (3.1-3.3) with $U = 0$ as $\Delta t \to 0$ [62]. Since under the condition of the theorem the SDE (3.1-3.3) is hypoelliptic it follows that $\mathscr{V}_T$ is positive definite. Now, since the set of invertible matrices is open in the topology corresponding to the standard matrix norms used here, there is a $\Delta t^* > 0$ so that $\mathscr{V}^{\lceil T/\Delta t \rceil}$ is invertible, hence positive definite for all $\Delta t \leq \Delta t^*$. In particular, for sufficiently small $\Delta t^* > 0$, the inequality (4.21) holds for

$$\underline{\mathscr{V}} = \mathscr{V}_T/2, \text{ and } \bar{\mathscr{V}} = 2\mathscr{V}_T. \tag{4.22}$$

The remainder of the proof is identical to [72]. Let Δt^* be sufficiently small so that (4.21) holds for the choice (4.22). For any set $E =\in \Omega_{\boldsymbol{q}} \times \Omega_{\boldsymbol{z}}$, we then have

$$\begin{aligned} &\mathbb{P}\left(\left(\mathscr{G}_{\boldsymbol{q}}{}^{\lfloor T/\Delta t \rfloor}, \mathscr{D}_{\boldsymbol{q}}{}^{\lfloor T/\Delta t \rfloor}\right) \in E\right) \\ &= \int_E (2\mu)^{2(n+m)} \det\left(\mathscr{V}^{\lfloor T/\Delta t \rfloor}\right)^{-1/2} \exp(-\frac{1}{2}\boldsymbol{x}^T \left(\mathscr{V}^{\lfloor T/\Delta t \rfloor}\right)^{-1} \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &\geq \int_E (2\mu)^{2(n+m)} 2^{-(n+m)} \det\left(\mathscr{V}_T\right)^{-1/2} \exp(-\boldsymbol{x}^T \mathscr{V}_T^{-1} \boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \end{aligned} \tag{4.23}$$

Now, define ν as

$$\nu(A_{\boldsymbol{q}} \times A_{\boldsymbol{z}}) = Z_R \inf_{\|\mathscr{D}_{\boldsymbol{q}}\|, \|\mathscr{D}_{\boldsymbol{z}}\| \leq R} \int_{A_{\boldsymbol{q}} - \mathscr{D}_{\boldsymbol{q}} \times A_{\boldsymbol{z}} - \mathscr{D}_{\boldsymbol{z}}} \exp(-\boldsymbol{x}^T \mathscr{V}^{-1} \boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{4.24}$$

The inequality (4.23) together with 2 implies that $\alpha = (2\mu)^{2(n+m)} 2^{-(n+m)} \det\left(\mathscr{V}\right)^{-1/2}$ ensures that (4.19) holds.

□

The following lemmas are required in the proofs of the above propositions.

[2]This is to be understood in terms of the corresponding quadratic forms i.e. for matrices A, B we say $A \leq B \iff \forall \boldsymbol{x} : \boldsymbol{x}^T A \boldsymbol{x} \leq \boldsymbol{x}^T B \boldsymbol{x}$,

Lemma 4.4.1. *$\boldsymbol{z}^{d_k} := \prod_{i=1}^{n+m} \boldsymbol{z}_i^{d_{k,i}}$ with $|d_k| := \sum_i d_{k,i} < 2l$ so that*

$$g_{\boldsymbol{q},l}(\boldsymbol{z}) := \sum_{k=1}^{K} f_k(\boldsymbol{q})\ \boldsymbol{z}^{d_k}$$

defines a polynomial in $\boldsymbol{z}$ of a degree less than $2l$ with coefficients $f_k \in \mathcal{C}^1(\Omega_{\boldsymbol{q}}, \mathbb{R})$, $\Omega_{\boldsymbol{q}}$ compact. There exists a constant $C_f > 0$ such that for all ϵ, $0 < \epsilon < 1$ and positive definite matrix $\boldsymbol{C} \in \mathbb{R}^{(n+m)\times(n+m)}$ with $\sigma_{\min}(\boldsymbol{C}) \geq 1$

$$\left| \sum_{k=1}^{K} f_k(\boldsymbol{q})\ \boldsymbol{z}^{d_k} \right| \leq K\epsilon(\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z})^l + \frac{1}{\epsilon^{2l-1}} C_f.$$

Proof. Let

$$f_{\max}(\boldsymbol{q}) := \max_{1\leq k\leq K} f_k(\boldsymbol{q}),$$

now,

$$\begin{aligned} g_{\boldsymbol{q},l}(\boldsymbol{z}) &\leq |f_{\max}(\boldsymbol{q})| \sum_{k=1}^{K} |\boldsymbol{z}^{d_k}| \\ &\leq \sum_{k=1}^{K} \epsilon^{2l-|d_k|} (\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z})^l + \frac{1}{\epsilon^{|d_k|}} |f_{\max}(\boldsymbol{q})| \end{aligned} \tag{4.25}$$

where $C_f := K\|f_{\max}\|_{L_\infty}$ and $\epsilon > 0$. In particular for $0 < \epsilon < 1$,

$$g_{\boldsymbol{q},l}(\boldsymbol{z}) \leq K\epsilon(\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z})^l + \frac{1}{\epsilon^{2l-1}} C_f.$$

□

Lemma 4.4.2. *Let $(\boldsymbol{Q}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}) \in M$, with M as defined in Section 3.1.1, $\boldsymbol{S}_{\Delta t}, \boldsymbol{F}_{\Delta t}$, as defined in (4.7), and $\boldsymbol{C}$ as symmetric positive definite matrix, such that*

$$\boldsymbol{\Gamma}\boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Gamma}^T$$

is positive definite, then

(i) $\boldsymbol{z}^T {\boldsymbol{F}_{\Delta t}}^T \boldsymbol{C} \boldsymbol{F}_{\Delta t} \boldsymbol{z} \leq \exp(-\Delta t\lambda) \boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z} \leq \exp(-\Delta t\lambda)\sigma_{\max}(\boldsymbol{C})\|\boldsymbol{z}\|_2^2$,
where $\lambda = \sigma_{\min}[\boldsymbol{\Gamma}^T\boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Gamma}](\sigma_{\max}(\boldsymbol{C}))^{-1}$,

(ii) $\|\sum_{k=0}^{\infty} {\boldsymbol{F}_{\Delta t}}^k\| < \infty$,

(iii) $\|\boldsymbol{F}_{\Delta t} - \boldsymbol{I}\| \leq \Delta t\|\boldsymbol{\Gamma}\| \exp(\|\boldsymbol{\Gamma}\|\Delta t) = O(\Delta t)$,

(iv) $\|\boldsymbol{S}_{\Delta t} {\boldsymbol{S}_{\Delta t}}^T\| \leq 2\Delta t \exp(\Delta t\|\boldsymbol{\Gamma}\|) c_1 \|\boldsymbol{Q}\|\|\boldsymbol{\Gamma}\| = O(\Delta t)$,

(v) $\|\boldsymbol{S}_{\Delta t}\| \leq (2\Delta t c_1 \exp(\Delta t\|\boldsymbol{\Gamma}\|)\|\boldsymbol{Q}\|\|\boldsymbol{\Gamma}\|)^{1/2} = O(\Delta t^{1/2})$,

(vi) $\|\boldsymbol{S}_{\Delta t}\boldsymbol{A}\boldsymbol{S}_{\Delta t}\| \leq 2\Delta t c_1 \exp(\Delta t\|\boldsymbol{\Gamma}\|)\|\boldsymbol{Q}\|\|\boldsymbol{\Gamma}\|\|\boldsymbol{A}\| = O(\Delta t)$,

where the constant $c_1 > 0$ is independent of Δt but may differ between the inequalities and may depend on the particular choice of matrix norm.

Proof. We note that $\|\cdot\|_{\boldsymbol{C}} : \mathbb{R}^n \to [0,\infty), \boldsymbol{z} \mapsto \boldsymbol{z}^T\boldsymbol{C}\boldsymbol{z}$ defines a Lyapunov function for the ODE $\dot{\boldsymbol{z}} = -\boldsymbol{\Gamma}\boldsymbol{z}$. In particular,

$$\begin{aligned}\partial_{\Delta t}\boldsymbol{z}^T{\boldsymbol{F}_{\Delta t}}^T\boldsymbol{C}\boldsymbol{F}_{\Delta t}\boldsymbol{z} &= -\boldsymbol{z}^T{\boldsymbol{F}_{\Delta t}}^T[\boldsymbol{\Gamma}^T\boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Gamma}]\boldsymbol{F}_{\Delta t}\boldsymbol{z} \\ &\leq -\frac{\sigma_{\min}[\boldsymbol{\Gamma}^T\boldsymbol{C} + \boldsymbol{C}\boldsymbol{\Gamma}]}{\sigma_{\max}(\boldsymbol{C})}\boldsymbol{z}^T{\boldsymbol{F}_{\Delta t}}^T\boldsymbol{C}\boldsymbol{F}_{\Delta t}\boldsymbol{z},\end{aligned}$$

which implies (i) by virtue of Grönwall's inequality. Statement (i) implies that $\boldsymbol{F}_{\Delta t}$ is a contraction with respect to the norm $\|\cdot\|_{\boldsymbol{C}}$, hence

$$\left\|\sum_{k=0}^{\infty}{\boldsymbol{F}_{\Delta t}}^k\right\| \leq \sum_{k=0}^{\infty}\left\|{\boldsymbol{F}_{\Delta t}}^k\right\|_{\boldsymbol{C}} = (1 - \|\boldsymbol{F}_{\Delta t}\|_{\boldsymbol{C}})^{-1}$$

which implies (ii). The inequality (iii) is obtained as a special case of the inequality

$$\|e^{A+B} - e^A\| \leq \|B\|e^{\|A\|}e^{\|B\|},$$

which holds for arbitrary real valued matrices A, B ([49], Corollary 6.2.32). We use (iii) to show (iv):

$$\begin{aligned}\|\boldsymbol{S}_{\Delta t}{\boldsymbol{S}_{\Delta t}}^T\| &= \|\boldsymbol{Q} - {\boldsymbol{F}_{\Delta t}}^T\boldsymbol{Q}\boldsymbol{F}_{\Delta t}\| \\ &= \|\boldsymbol{Q}(\boldsymbol{I} - \boldsymbol{F}_{\Delta t}) + (\boldsymbol{I} - {\boldsymbol{F}_{\Delta t}}^T)\boldsymbol{Q}\boldsymbol{F}_{\Delta t}\| \\ &\leq \|\boldsymbol{Q}\|\|\boldsymbol{I} - \boldsymbol{F}_{\Delta t}\|\|\boldsymbol{I} + \boldsymbol{F}_{\Delta t}\| \\ &\leq \Delta t\ \|\boldsymbol{I} + \boldsymbol{F}_{\Delta t}\|\|\boldsymbol{Q}\|\|\boldsymbol{\Gamma}\|\exp(\|\boldsymbol{\Gamma}\|\Delta t) \\ &\leq \Delta t 2\|\boldsymbol{Q}\|\|\boldsymbol{\Gamma}\|\exp(\|\boldsymbol{\Gamma}\|\Delta t).\end{aligned}$$

For the particular choice of the Frobenius norm we have

$$\|\boldsymbol{S}_{\Delta t}\|_{\mathrm{F}}^2 = \|\boldsymbol{S}_{\Delta t}{\boldsymbol{S}_{\Delta t}}^T\|_{\mathrm{F}},$$

hence

$$\|\boldsymbol{S}_{\Delta t}\|_{\mathrm{F}} \leq \left(\Delta t c_1 \exp(\Delta t\|\boldsymbol{\Gamma}\|_{\mathrm{F}})\|\boldsymbol{Q}\|_{\mathrm{F}}\|\boldsymbol{\Gamma}\|_{\mathrm{F}}\right)^{1/2},$$

and (v) for general matrix norms follows from the equality of norms in finite dimension. Inequality (vi) follows directly from (v) and the submultiplicative property of matrix norms. □

Remark 4.4.1. *In a proof for a symmetric spitting scheme one would use the same decomposition as used in (4.18) to show the uniform Lyapunov condition. While the first term in the decomposition would be identical, the remainder term would be comprised of slightly different (and more) terms. The respective additional terms are of lower polynomial order than the first term in the second line in (4.18) an thus can be bounded using the same estimates as employed in the proof presented here. Similarly, the proof of the uniform Lyapunov condition would rely on the same decomposition into a non-degenerated Gaussian terms and a bounded term as used in the proof presented here.*

4.5 White noise and overdamped limit for for H-OU splitting methods

In this section we show that integrators which are constructed as symmetric H-OU splitting schemes for Markovian reformulation (3.1-3.3) of GLEs with constant diffusion and dissipation terms are consistent with the asymptotic limits derived in Section 3.5. That is, we show that both in the overdamped limit (OD-limit) and in the white noise limit (WN-limit), these integrators reduce to consistent integration schemes for the underdamped Langevin equation and the overdamped Langevin equation, respectively. Most importantly, both the (weak) order of the discretisation error and the maximum admissible stepsize Δt^* are not affected in the respective limits. From a practical viewpoint this is of great importance, as this means that H-OU splitting schemes can be seen as universal integration schemes for the overdamped Langevin equation, the underdamped Langevin equation and the generalised Langevin equation.

4.5.1 White noise limit

Recall that in the view of the rescaled process (3.64) the white noise limit (WN-limit) corresponds to a rescaling of the matrix $\mathbf{\Gamma}$ as

$$\mathbf{\Gamma}^\epsilon = \begin{pmatrix} \mathbf{\Gamma}_{1,1} & \epsilon^{-1}\mathbf{\Gamma}_{1,2} \\ \epsilon^{-1}\mathbf{\Gamma}_{2,1} & \epsilon^{-2}\mathbf{\Gamma}_{2,2} \end{pmatrix}. \tag{4.26}$$

Recall also that in H-OU splitting schemes the matrix $\mathbf{\Gamma}$ appears in the integration of the O-part (and only therein). For rescaled the process (3.64) the stochastic flow-map associated with the O-part is of the form

$$\Phi^O_{\Delta t} : (\boldsymbol{q}, \boldsymbol{p}, \boldsymbol{s}) \mapsto (\boldsymbol{q}, \boldsymbol{F}^\epsilon_{\Delta t}\,(\boldsymbol{p}, \boldsymbol{s})^T + \boldsymbol{S}^\epsilon_{\Delta t}\mathscr{R}),\ \mathscr{R} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{n+m}), \tag{4.27}$$

where

$$\boldsymbol{F}^\epsilon_{\Delta t} := \exp(-\Delta t \mathbf{\Gamma}^\epsilon),\ \Delta t > 0,$$

and $\boldsymbol{S}^\epsilon_{\Delta t}$ is a matrix square root of

$$\begin{pmatrix} \boldsymbol{I}_n & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q} \end{pmatrix} - \boldsymbol{F}^\epsilon_{\Delta t} \begin{pmatrix} \boldsymbol{I}_n & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q} \end{pmatrix} \boldsymbol{F}^{\epsilon T}_{\Delta t},$$

with $\boldsymbol{Q}$ as defined in Proposition 3.1.1. Therefore, the limit

$$\exp\left(-\Delta t \begin{pmatrix} \mathbf{\Gamma}_{1,1} & \epsilon^{-1}\mathbf{\Gamma}_{1,2} \\ \epsilon^{-1}\mathbf{\Gamma}_{2,1} & \epsilon^{-2}\mathbf{\Gamma}_{2,2} \end{pmatrix}\right),\ \text{as } \epsilon \to 0, \tag{4.28}$$

is therefore central for the understanding of properties of H-OU integrators in the white noise limit. In the following Proposition 4.5.1 we prove convergence $\boldsymbol{F}^\epsilon_{\Delta t}$ to a singular matrix under additional assumption on the structure of $\mathbf{\Gamma}$. In result 4.5.1 we present a formal derivation of the limit (4.28) for more general forms of $\mathbf{\Gamma}$.

Proposition 4.5.1. *Let*

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{0} & -\boldsymbol{D}_a \\ \boldsymbol{D}_a & \boldsymbol{D}_b \end{pmatrix} \in \mathbb{R}^{2n\times 2n}$$

with $\boldsymbol{D}_a := \mathrm{diag}(a_1, a_2, \dots a_n)$, $\boldsymbol{D}_b := \mathrm{diag}(b_1, b_2, \dots b_n)$, *where the entries* $a_i, b_i,\ 1 \leq$

$i \leq n$ are positive scalars. For $\mathbf{\Gamma}^\epsilon$ as specified in (4.26) we find

$$\lim_{\epsilon\to 0} \exp(-\Delta t \mathbf{\Gamma}^\epsilon) = \begin{pmatrix} \exp(-\Delta t \boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Proof. Without loss of generality let $\Delta t = 1$. Let further

$$\boldsymbol{O} = \hat{\boldsymbol{I}}^{(2n)}_{2n-1,2n} \hat{\boldsymbol{I}}^{(2n)}_{2n-3,2n-1} \dots \hat{\boldsymbol{I}}^{(2n)}_{5,n+3} \hat{\boldsymbol{I}}^{(2n)}_{3,n+2} \hat{\boldsymbol{I}}^{(2n)}_{1,n+1},$$

where $\hat{\boldsymbol{I}}^{(2n)}_{i,j}$ denotes the elementary matrix whose action when multiplied from the left to a matrix $\boldsymbol{A} \in \mathbb{R}^{2n\times 2n}$ corresponds to a swap of i-th and j-th rows of $\boldsymbol{A}$. A simple calculation shows that the matrix $\mathbf{\Gamma}^\epsilon$ can be written in terms of the following orthogonal similarity transformation as

$$\mathbf{\Gamma}^\epsilon = \boldsymbol{O}^T \begin{pmatrix} A_1^\epsilon & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & A_2^\epsilon & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & A_n^\epsilon \end{pmatrix} \boldsymbol{O}$$

with

$$A_i^\epsilon = \begin{pmatrix} 0 & -\epsilon^{-1} a_i \\ \epsilon^{-1} a_i & \epsilon^{-2} b_i \end{pmatrix},\ 1 \leq i \leq n.$$

Since also

$$\boldsymbol{O}^T \begin{pmatrix} \exp(-\Delta t \boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{O} = \begin{pmatrix} \exp(-\Delta t \boldsymbol{D}_a^2 \boldsymbol{D}_b^{-1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

it is therefore sufficient to show the result for a 2×2 matrix,

$$A^\epsilon = \begin{pmatrix} 0 & -\epsilon^{-1} a \\ \epsilon^{-1} a & \epsilon^{-2} b \end{pmatrix}$$

with $a, b > 0$: Let $c_\epsilon = \sqrt{b^2 - 4a^2\epsilon^2}$. The eigenvalues of the matrix A^ϵ are

$$\lambda_1^\epsilon = \frac{c_\epsilon + b}{2\epsilon^2},\ \lambda_2^\epsilon = \frac{-c_\epsilon + b}{2\epsilon^2},$$

with corresponding eigenvectors

$$v_1 = \begin{pmatrix} -\frac{b-c_\epsilon}{2a\epsilon} \\ 1 \end{pmatrix},\ v_2 = \begin{pmatrix} -\frac{c_\epsilon+b}{2a\epsilon} \\ 1 \end{pmatrix},$$

respectively, thus

$$\begin{aligned} \exp(A^\epsilon) &= [v_1, v_2] \mathrm{diag}(e^{-\lambda_1^\epsilon}, e^{-\lambda_2^\epsilon}) [v_1, v_2]^{-1} \\ &= \begin{pmatrix} e^{-\lambda_2^\epsilon} \frac{(b+c_\epsilon)}{2c_\epsilon} - e^{-\lambda_1^\epsilon} \frac{(b-c_\epsilon)}{2c_\epsilon} & \left(e^{-\lambda_2^\epsilon} - e^{-\lambda_1^\epsilon}\right) \frac{a\epsilon}{c_\epsilon} \\ -\left(e^{-\lambda_2^\epsilon} - e^{-\lambda_1^\epsilon}\right) \frac{a\epsilon}{c_\epsilon} & \frac{e^{-\lambda_1^\epsilon}(b+c_\epsilon)}{2c_\epsilon} - \frac{e^{-\lambda_2^\epsilon}(b-c_\epsilon)}{2c_\epsilon} \end{pmatrix}, \end{aligned}$$

and the result follows since

$$\lambda_1^\epsilon \to \infty,\ \lambda_2^\epsilon \to \frac{a^2}{b},\ c_\epsilon \to b,$$

as $\epsilon \to 0$. $\square$

Result 4.5.1. *Let*

$$\mathbf{\Gamma}^\epsilon = \begin{pmatrix} \mathbf{\Gamma}_{1,1} & \epsilon^{-1}\mathbf{\Gamma}_{1,2} \\ \epsilon^{-1}\mathbf{\Gamma}_{2,1} & \epsilon^{-2}\mathbf{\Gamma}_{2,2} \end{pmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}, \epsilon > 0,$$

such that $-\mathbf{\Gamma}$ *is stable. Let for* $\Delta t > 0$,

$$\boldsymbol{F}^\epsilon_{\Delta t} = \exp(-\Delta t \mathbf{\Gamma}^\epsilon).$$

As $\epsilon \to 0$ *the matrix exponential* $\boldsymbol{F}^\epsilon_{\Delta t}$ *converges to the singular matrix*

$$\begin{pmatrix} e^{-\Delta t \hat{\mathbf{\Gamma}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

where $\hat{\mathbf{\Gamma}} = \mathbf{\Gamma}_{1,1} - \mathbf{\Gamma}_{1,2}\mathbf{\Gamma}_{2,2}^{-1}\mathbf{\Gamma}_{2,1}$.

Derivation. We use a singular perturbation approach to show this result: Let

$$\begin{aligned} \frac{\partial u}{\partial t} &= -\mathbf{\Gamma}^\epsilon u \\ &= -\begin{pmatrix} \mathbf{\Gamma}_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} u - \frac{1}{\epsilon}\begin{pmatrix} \mathbf{0} & \mathbf{\Gamma}_{1,2} \\ \mathbf{\Gamma}_{2,1} & \mathbf{0} \end{pmatrix} u - \frac{1}{\epsilon^2}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{2,2} \end{pmatrix} u \end{aligned} \tag{4.29}$$

with $u = (v_1^T, v_2^T)^T$, where $v_1(t) \in \mathbb{R}^n$ and $v_2(t) \in \mathbb{R}^m$. Formally expanding u in ϵ, so that $u = \sum_{i=0}^\infty u_i \epsilon^i$ and equating (4.29) in powers of ϵ yields the following sequence of equations

$$-\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{2,2} \end{pmatrix} u_0 = \mathbf{0} \tag{4.30}$$

$$-\begin{pmatrix} \mathbf{0} & \mathbf{\Gamma}_{1,2} \\ \mathbf{\Gamma}_{2,1} & \mathbf{0} \end{pmatrix} u_0 - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{2,2} \end{pmatrix} u_1 = \mathbf{0} \tag{4.31}$$

$$-\begin{pmatrix} \mathbf{\Gamma}_{1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} u_0 - \begin{pmatrix} \mathbf{0} & \mathbf{\Gamma}_{1,2} \\ \mathbf{\Gamma}_{2,1} & \mathbf{0} \end{pmatrix} u_1 - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{2,2} \end{pmatrix} u_2 = \frac{\partial u_0}{\partial t} \tag{4.32}$$

Let

$$u_i^T = (v_{i,1}^T, v_{i,2}^T)^T, 1 \leq i,$$

with $v_{i,1}(t) \in \mathbb{R}^n, v_{i,2}(t) \in \mathbb{R}^m$. From equation (4.30) it follows that $v_{i,2} = \mathbf{0}$, hence $u_0(t) = \begin{pmatrix} v_{0,1}(t) \\ \mathbf{0} \end{pmatrix}$, where $v_{0,1}$ is a real valued function. Let $v_1(t) := u_1(t) - \sum_{i=1}^n \boldsymbol{e}_i \langle \boldsymbol{e}_1, u_1(t) \rangle$. From equation (4.31) we obtain

$$\mathbf{\Gamma}_{2,1} v_0(t) - \mathbf{\Gamma}_{2,2} v_1(t) = \mathbf{0},$$

hence

$$v_1(t) = \mathbf{\Gamma}_{2,2}^{-1}\mathbf{\Gamma}_{2,1} v_0(t).$$

Finally, form equation (4.32) it follows that

$$\frac{\partial v_0(t)}{\partial t} = -\mathbf{\Gamma}_{2,1} v_1(t),$$

and therefore

$$\frac{\partial v_0(t)}{\partial t} = \mathbf{\Gamma}_{1,1} - \mathbf{\Gamma}_{1,2}\mathbf{\Gamma}_{2,2}^{-1}\mathbf{\Gamma}_{2,1}v_0(t).$$

□

The following corollary 4.5.1 and Remark 4.5.1 restate the above derived results in the context of the numerical integration scheme.

Collorary 4.5.1 (White noise limit for H-OU splitting schemes)**.** *Consider the scaling*

$$\lambda = 1,\ \mu_1 = 0,\ \mu_2 = \epsilon^{-1},\ \mu_3 = \epsilon^{-2}.$$

and let $\mathbf{\Gamma}$ *be as specified in Proposition 4.5.1. Let* $\hat{\Phi}^{\mathrm{XYZYX}}_{\Delta t,\epsilon}$ *be a H-OU splitting scheme applied to the rescaled GLE (3.64). In the asymptotic limit* $\epsilon \to 0$ *the integration scheme* $\hat{\Phi}^{\mathrm{XYZYX}}_{\Delta t,\epsilon}$ *coincides in distribution with the corresponding splitting scheme* $\hat{\Phi}^{\mathrm{XYZYX}}_{\Delta t,\mathrm{LD}}$ *for the underdamped Langevin equation,*

$$\begin{aligned} \mathrm{d}\hat{\boldsymbol{q}} &= \hat{\boldsymbol{p}}, \\ \mathrm{d}\hat{\boldsymbol{p}} &= -\nabla U(\hat{\boldsymbol{q}}) - \hat{\mathbf{\Gamma}}\hat{\boldsymbol{p}} + \beta^{-1/2}\hat{\mathbf{\Sigma}}\mathrm{d}\boldsymbol{W}_1, \end{aligned} \tag{4.33}$$

with friction tensor

$$\hat{\mathbf{\Gamma}} = \boldsymbol{D}_a^2\boldsymbol{D}_b^{-1},$$

and diffusion tensor

$$\hat{\mathbf{\Sigma}} = \sqrt{2}\boldsymbol{D}_a\boldsymbol{D}_b^{-1/2}.$$

More precisely,

$$\lim_{\epsilon\to 0}\mathbb{E}_{\mathscr{R}}\left[\varphi\left(\hat{\Phi}^{\mathrm{XYZYX}}_{\Delta t,\epsilon}(x)\right)\right] = \mathbb{E}_{\mathscr{R}}\left[\varphi\left(\hat{\Phi}^{\mathrm{XYZYX}}_{\Delta t,\mathrm{LD}}(q,p)\right)\right]$$

for all $\varphi \in \mathcal{C}_b^\infty(\Omega_{\boldsymbol{q}} \times \Omega_{\boldsymbol{p}}),\ x = (q,p,s) \in \Omega_{\boldsymbol{x}}$*, where* $\mathbb{E}_{\mathscr{R}}[\,\cdot\,]$ *denotes the expectation with respect to* $\mathscr{R}$ *in the stochastic flow map* (4.27)*.*

Remark 4.5.1. *Let* $-\mathbf{\Gamma} \in \mathbb{R}^{(n+m)\times(n+m)}$ *be stable. Under the assumption that the expansion in the derivation of result 4.5.1 is valid, the statement of corollary 4.5.1 is valid with*

$$\hat{\mathbf{\Gamma}} = \mathbf{\Gamma}_{1,1} - \mathbf{\Gamma}_{2,1}\mathbf{\Gamma}_{2,2}^{-1}\mathbf{\Gamma}_{1,2},$$

and $\hat{\mathbf{\Sigma}}$ *such that*

$$\hat{\mathbf{\Sigma}}\hat{\mathbf{\Sigma}}^T = \hat{\mathbf{\Gamma}} + \hat{\mathbf{\Gamma}}^T.$$

4.5.2 Overdamped limit

The derivation of the overdamped limit is of the Markovian reformulation (3.1-3.3) of the GLE is much simpler than the white-noise limit. Depending on the integration order one recovers either an Euler-Maruyama scheme or the Leimkuhler-Matthews method in the asymptotic limit $\epsilon \to 0$.

Proposition 4.5.2. *Consider the scaling*

$$\lambda = 1,\ \mu_1 = \epsilon^{-1},\ \mu_2 = \epsilon^{-1},\ \mu_3 = \epsilon^{-1}.$$

and let $\hat{\Phi}^{\mathrm{XYZYX}}_{\Delta t,\epsilon}$ *be an H-OU splitting scheme applied to the rescaled GLE (3.64). Let X=B, Y=A, Z=O. In the asymptotic limit* $\epsilon \to 0$ *the respective integration scheme*

coincides in law with the Leimkuhler-Matthews method [70],

$$\boldsymbol{q}_{k+1} = \boldsymbol{q}_k - \widetilde{\Delta t}\nabla_{\boldsymbol{q}} U(\boldsymbol{q}_k) + \sqrt{2\widetilde{\Delta t}\beta^{-1}}\frac{\widetilde{\mathscr{R}}_k + \widetilde{\mathscr{R}}_{k-1}}{2}, \tag{4.34}$$

with $\widetilde{\Delta t} = \frac{\Delta t^2}{2}$ *and* $\widetilde{\mathscr{R}}^k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n), k \in \mathbb{N}$ *i.i.d. Similarly, let X=O, Y=B, Z=A. In the asymptotic limit* $\epsilon \to 0$ *the respective integration scheme coincides in law with the Euler-Maruyama method*

$$\boldsymbol{q}_{k+1} = \boldsymbol{q}_k - \widetilde{\Delta t}\nabla_{\boldsymbol{q}} U(\boldsymbol{q}_k) + \sqrt{2\widetilde{\Delta t}\beta^{-1}}\mathscr{R}_k,$$

with $\widetilde{\Delta t} = \frac{\Delta t^2}{2}$ *and* $\widetilde{\mathscr{R}}^k$ *as above.*

Proof. We have $\boldsymbol{F}_{\Delta t} = \exp(-\frac{\Delta t}{\epsilon}\boldsymbol{\Gamma}) \to \mathbf{0}$ as $\epsilon \to 0$, and therefore also

$$\boldsymbol{S}_{\Delta t} \to \beta^{-1/2}\boldsymbol{I}_{n+m}, \text{ as } \epsilon \to 0.$$

Consequently, in the limit $\epsilon \to 0$ the updates of $\boldsymbol{p}_k$ (and of course also $\boldsymbol{q}_k$) are not affected by the value of $\boldsymbol{s}_k$. In the case of gle-BAOAB the effective dynamics of $\boldsymbol{q}_k$ and $\boldsymbol{p}_k$ are then described by the updating sequence

$$\begin{aligned}
\boldsymbol{p}_{k+1/2} &= \boldsymbol{p}_k - \frac{\Delta t}{2}\nabla U(\boldsymbol{q}_k),\\
\boldsymbol{q}_{k+1/2} &= \boldsymbol{q}_k + \frac{\Delta t}{2}\boldsymbol{p}_{k+1/2},\\
\hat{\boldsymbol{p}}_{k+1/2} &= \sqrt{\beta^{-1}}\tilde{\mathscr{R}}_k,\\
\boldsymbol{q}_{k+1} &= \boldsymbol{q}_{k+1/2} + \frac{\Delta t}{2}\hat{\boldsymbol{p}}_{k+1/2},\\
\boldsymbol{p}_{k+1} &= \hat{\boldsymbol{p}}_{k+1/2} - \frac{\Delta t}{2}\nabla U(\boldsymbol{q}_{k+1}),
\end{aligned} \tag{4.35}$$

from which the momentum component can be eliminated, resulting in the recursion (4.34) for the position component. The result for gle-OBABO follows in the same way. □

4.6 Error analysis for ergodic averages

In this section we provide results regarding the convergence order in Δt of the discretisation bias in ergodic averages of symmetric H-OU splitting schemes. We first derive the explicit form of the method and Δt dependent invariant measure in the case of a linear force. Secondly, we formally show for general potential functions that the gle-BAOAB scheme possesses a super-convergence property in the discrete time version of the overdamped limit (OD-limit), i.e., the discretisation bias in ergodic averages of observables which are purely functions of the configurational variable $\boldsymbol{q}$ is $O(\Delta t^4)$; (instead of $O(\Delta t^2)$ as one would expect by construction).

4.6.1 Error analysis for a quadratic potential

In this section we analyze the sampling properties of the symmetric H-OU splitting schemes for a system whose force is derived from an harmonic potential of the form $U(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T\boldsymbol{\Omega}\boldsymbol{q}$. The main result is summarised in the following proposition.

Proposition 4.6.1. *Let $U(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^T\boldsymbol{\Omega}\,\boldsymbol{q}$ with $\boldsymbol{\Omega} \in \mathbb{R}^{n\times n}$ positive definite. The Gaussian measure with associated probability density $\hat{\rho}_\beta(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, with*

$$\boldsymbol{\mu}_\beta = \mathbf{0}, \quad \boldsymbol{\Sigma}_\beta = \beta^{-1}\mathrm{diag}\left(\boldsymbol{\Omega}^{-1},\ \left(1-\Delta t^2/4\right)\boldsymbol{M},\ \boldsymbol{Q}\right), \tag{4.36}$$

is invariant under $\hat{\Phi}^{\mathrm{BAOAB}}_{\Delta t}$ and $\hat{\Phi}^{\mathrm{ABOBA}}_{\Delta t}$. Similarly, the stochastic flow maps $\hat{\Phi}^{\mathrm{OBABO}}_{\Delta t}$ and $\hat{\Phi}^{\mathrm{OABAO}}_{\Delta t}$ preserve the gaussian measure with the probability density $\widehat{\rho}(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, with

$$\boldsymbol{\mu}_\beta = \mathbf{0}, \quad \boldsymbol{\Sigma}_\beta = \beta^{-1}\mathrm{diag}\left(\left(1-\Delta t^2/4\right)\boldsymbol{\Omega}^{-1},\ \boldsymbol{M},\ \boldsymbol{Q}\right)$$

with $\boldsymbol{Q}$ as defined in Proposition 3.1.1.

Proof. In order to prove the invariance of the measures under the corresponding evolution operator we need to show in each case that the equation

$$\hat{\rho}_\beta = \exp(\frac{\Delta t}{2}\mathcal{L}_X^\dagger)\exp(\frac{\Delta t}{2}\mathcal{L}_Y^\dagger)\exp(\Delta t\mathcal{L}_Z^\dagger)\exp(\frac{\Delta t}{2}\mathcal{L}_Y^\dagger)\exp(\frac{\Delta t}{2}\mathcal{L}_X^\dagger)\hat{\rho}_\beta, \tag{4.37}$$

with $\hat{\rho}_\beta(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x};\ \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, holds.
For a linear force $-\nabla U(\boldsymbol{q}) = -\boldsymbol{\Omega}\,\boldsymbol{q}$ and fixed time step Δt, the flow maps $\Phi^A_{\Delta t}$ and $\Phi^B_{\Delta t}$ reduce to linear transformations

$$\begin{aligned}\Phi^A_{\Delta t}(\boldsymbol{x}) &= \Psi_A\boldsymbol{x},\\ \Phi^B_{\Delta t}(\boldsymbol{x}) &= \Psi_B\boldsymbol{x},\end{aligned} \tag{4.38}$$

where

$$\boldsymbol{\Psi}_A = \begin{bmatrix}\boldsymbol{I}_n & \Delta t\boldsymbol{I}_n & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I}_m\end{bmatrix}, \quad \boldsymbol{\Psi}_B = \begin{bmatrix}\boldsymbol{I}_n & \mathbf{0} & \mathbf{0}\\ -\Delta t\boldsymbol{\Omega} & \boldsymbol{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{I}_m\end{bmatrix}.$$

Consequently, the action of the forward operators $\exp(\Delta t\mathcal{L}_A^\dagger)$ and $\exp(\Delta t\mathcal{L}_B^\dagger)$ on a gaussian density is explicitly given as

$$\exp(\Delta t\mathcal{L}_A^\dagger)\mathcal{N}(\ \cdot\ |\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\ \cdot\ |\ \boldsymbol{\Psi}_A\boldsymbol{\mu}, \boldsymbol{\Psi}_A\boldsymbol{\Sigma}\boldsymbol{\Psi}_A^T), \tag{4.39}$$

$$\exp(\Delta t\mathcal{L}_B^\dagger)\mathcal{N}(\ \cdot\ |\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\ \cdot\ |\ \boldsymbol{\Psi}_B\boldsymbol{\mu}, \boldsymbol{\Psi}_B\boldsymbol{\Sigma}\boldsymbol{\Psi}_B^T). \tag{4.40}$$

And similarly one finds for the OU-part:

$$\exp(\Delta t\mathcal{L}_O^\dagger)\mathcal{N}(\ \cdot\ |\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}(\ \cdot\ |\boldsymbol{F}_{\Delta t}\boldsymbol{\mu}, \boldsymbol{F}_{\Delta t}\boldsymbol{\Sigma}{\boldsymbol{F}_{\Delta t}}^T + \boldsymbol{S}_{\Delta t}{\boldsymbol{S}_{\Delta t}}^T). \tag{4.41}$$

Since the forward propagators conserve the Gaussian structure of the density, the equation (4.37) is well posed and we can use the relation (4.39-4.41) to derive the invariance condition for the parameters $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\Sigma}_\beta$ for each integration sequence. In the case of gle-BAOAB, we find

$$\begin{aligned}\boldsymbol{\mu}_\beta &= \boldsymbol{\Psi}_B\boldsymbol{\Psi}_A\boldsymbol{F}_{\Delta t}\boldsymbol{\Psi}_A\boldsymbol{\Psi}_B\boldsymbol{\mu}_\beta,\\ \boldsymbol{\Sigma}_\beta &= \boldsymbol{\Psi}_B\boldsymbol{\Psi}_A\boldsymbol{F}_{\Delta t}\boldsymbol{\Psi}_A\boldsymbol{\Psi}_B\boldsymbol{\Sigma}_\beta\boldsymbol{\Psi}_B^T\boldsymbol{\Psi}_A^T{\boldsymbol{F}_{\Delta t}}^T\boldsymbol{\Psi}_A^T\boldsymbol{\Psi}_B^T + \boldsymbol{\Psi}_B\boldsymbol{\Psi}_A\boldsymbol{S}_{\Delta t}{\boldsymbol{S}_{\Delta t}}^T\boldsymbol{\Psi}_A^T\boldsymbol{\Psi}_B^T.\end{aligned} \tag{4.42}$$

One directly verifies that the parameter choice (4.36) solves this system of linear equations. The statements for gle-ABOBA, gle-OBABO, and gle-OABAO, are shown in an analogous way. □

4.6.2 Super-convergence of gle-BAOAB

Let

$$\mathbf{\Gamma} = \begin{pmatrix} \Gamma_{1,1} & \Gamma_{1,2} \\ \Gamma_{2,1} & \Gamma_{2,2} \end{pmatrix} \in \mathbb{R}^2,$$

and $U \in \mathcal{C}^\infty(\mathbb{T}^d, \mathbb{R})$ such that the SDE (3.1-3.3) is ergodic. In this section we provide a generalisation of a super-convergence result which has been previously derived in [70] for a BAOAB discretisation of the underdamped Langevin equation; (see also [72] for a rigorous proof). For this purpose we consider the scaling

$$\lambda = 1, \mu_1 = \mu_2 = \mu_3 = \epsilon^{-1}, \tag{4.43}$$

which up to the missing rescaling of the time variable corresponds to the overdamped limit defined in Section 3.5. We formally show that as $\epsilon \to 0$, the leading error term in the discretisation bias of ergodic averages of observables which are constant in $\boldsymbol{p}, \boldsymbol{s}$, decreases linearly in ϵ, i.e.,

$$\left| \int_{\Omega_{\boldsymbol{x}}} \varphi(\boldsymbol{q}) \mu_{\Delta t}(\mathrm{d}\boldsymbol{x}) - \mathbb{E}_\mu \varphi \right| = O(\epsilon \Delta t^2), \tag{4.44}$$

which in the asymptotic limit results in an observed 4th order convergence of the discretisation bias, i.e.,

$$\lim_{\epsilon \to 0} \left| \int_{\Omega_{\boldsymbol{x}}} \varphi(\boldsymbol{q}) \mu_{\Delta t}(\mathrm{d}\boldsymbol{x}) - \mathbb{E}_\mu \varphi \right| = O(\Delta t^4). \tag{4.45}$$

While in the view of Proposition 4.5.2 the super-convergence in the asymptotic limit is not surprising, the linear decrease of the pre-constant in the leading error term for ergodic averages for sufficiently small $\epsilon > 0$ provides additional insight in the dependency of the invariant measure $\mu_{\Delta t}$ on $\mathbf{\Gamma}$.

Let $\mathcal{L}^\dagger_{\mathrm{GLE}}$ be the Fokker-Planck operator of the exact GLE dynamics and $\rho_{\boldsymbol{Q},\beta}$ the solution of the corresponding stationary equation, i.e.,

$$\mathcal{L}^\dagger_{\mathrm{GLE}} \rho_{\boldsymbol{Q},\beta} = 0,$$

Furthermore, we formally define $\widehat{\mathcal{L}}^\dagger_{\text{gle-BAOAB}}$ via the relation

$$e^{t\widehat{\mathcal{L}}^\dagger_{\text{gle-BAOAB}}} = \hat{\mathcal{P}}^{\dagger \mathrm{BAOAB}}_{\Delta t},$$

Let Δt^* be as defined in Proposition 4.4.1 so that for all $\Delta t, 0 < \Delta t < \Delta t^*$ the Markov chain generated by the numerical scheme is ergodic with invariant density $\hat{\rho}_{\Delta t}$, i.e.,

$$\widehat{\mathcal{L}}^\dagger_{\text{gle-BAOAB}} \hat{\rho}_{\Delta t} = 0. \tag{4.46}$$

For the scaling (4.43), the Fokker-Planck operator associated with the rescaled version of the GLE as defined in (3.64), can be decomposed as

$$\mathcal{L}^\dagger_{\mathrm{GLE}} = \mathcal{L}^\dagger_H + \frac{1}{\epsilon} \mathcal{L}^\dagger_O,$$

where $\mathcal{L}^\dagger_O$ is the Fokker-Planck operator corresponding to the O-part of the unscaled

process (3.1-3.3).
We consider a formal expansion of the operator $\widehat{\mathcal{L}}^\dagger_{\text{gle-BAOAB}}$ via the BCH formula (see Section 2.8.1), i.e.,

$$\widehat{\mathcal{L}}^\dagger_{\text{gle-BAOAB}} = \mathcal{L}_{\text{GLE}} + \Delta t^2 \mathcal{L}_2^\dagger + \Delta t^4 \mathcal{L}_4^\dagger + \dots,$$

where the leading term $\mathcal{L}_2^\dagger$ is of the form

$$\begin{aligned}\mathcal{L}_2^\dagger =& \frac{1}{12}\left([\mathcal{L}_O^\dagger,[\mathcal{L}_O^\dagger,\mathcal{L}_A^\dagger]] + [\mathcal{L}_O^\dagger + \mathcal{L}_A^\dagger,[\mathcal{L}_O^\dagger + \mathcal{L}_A^\dagger,\mathcal{L}_B^\dagger]]\right) \\ &-\frac{1}{24}\left([\mathcal{L}_A^\dagger,[\mathcal{L}_A^\dagger,\mathcal{L}_O^\dagger]] + [\mathcal{L}_B^\dagger,[\mathcal{L}_B^\dagger,\mathcal{L}_A^\dagger + \mathcal{L}_B^\dagger]]\right).\end{aligned} \tag{4.47}$$

Similarly, we assume that $\hat{\rho}_{\Delta t}$ can be expanded in even powers of Δt so that

$$\hat{\rho}_{\Delta t} = \rho_{\boldsymbol{Q},\beta}\left(1 + \sum_{i=1}^{\infty} \Delta t^{2i} f_i\right)$$

with

$$\int_{\Omega_{\boldsymbol{x}}} f_i(\boldsymbol{x})\rho_{\boldsymbol{Q},\beta}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 0.$$

Plugging this expression into (4.46),

$$\widehat{\mathcal{L}}^\dagger_{\text{gle-BAOAB}}\left(\rho_{\boldsymbol{Q},\beta}\left(1 + \sum_{i=1}^{\infty} \Delta t^{2i} f_i\right)\right) = 0, \tag{4.48}$$

and equating in powers of Δt, we obtain a collection of PDEs whose solutions correspond to the correction terms $f_{2i},\ i \geq 1$. For the leading order correction term f_2, we find

$$\epsilon^{-1}\mathcal{L}_O^\dagger(\rho_{\boldsymbol{Q},\beta} f_2) + \mathcal{L}_H^\dagger(\rho_{\boldsymbol{Q},\beta} f_2) = \mathcal{L}_2^\dagger(\rho_{\boldsymbol{Q},\beta}), \tag{4.49}$$

with

$$\begin{aligned}\mathcal{L}_2^\dagger\rho_{\boldsymbol{Q},\beta} = \Big[&-\frac{1}{4}\beta\left(\boldsymbol{p}^2\Gamma_{1,1} + \boldsymbol{p}\,\boldsymbol{s}\Gamma_{2,1} - \beta^{-1}\Gamma_{1,1}\right)U''(\boldsymbol{q}) \\ &-\frac{1}{12}\beta\boldsymbol{p}^3U^{(3)}(\boldsymbol{q}) + \frac{1}{4}\beta\boldsymbol{p}U'(\boldsymbol{q})U''(\boldsymbol{q})\Big]\rho_{\boldsymbol{Q},\beta}.\end{aligned} \tag{4.50}$$

Since $\rho_{\boldsymbol{Q},\beta}$ is positive (in particular, the reciprocal $1/\rho_{\boldsymbol{Q},\beta}$ is well defined), there exists a solution for (4.49), if and only if there exists a solution ψ for the equation

$$\left(\epsilon^{-1}\mathcal{L}_O^\dagger + \mathcal{L}_H^\dagger\right)\psi = \mathcal{L}_2(\rho_{\boldsymbol{Q},\beta}). \tag{4.51}$$

By virtue of the Fredholm alternative it follows that (4.51) possesses a solution if and only if all functions g which lie in the null space of the adjoint of $\left(\epsilon^{-1}\mathcal{L}_O^\dagger + \mathcal{L}_H^\dagger\right)$ satisfy $\langle g, \mathcal{L}_2\rho_{\boldsymbol{Q},\beta}\rangle = 0$, where $\langle\cdot,\cdot\rangle$ denotes the standard L^2 scalar product. In other words, there exists a solution ψ for (4.51) iff the implication

$$\left(\epsilon^{-1}\mathcal{L}_O + \mathcal{L}_H\right)g = 0 \Rightarrow \int_{\Omega_{\boldsymbol{x}}} g(\boldsymbol{x})\mathcal{L}_2\rho_{\boldsymbol{Q},\beta}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 0, \tag{4.52}$$

is true. Since the SDE associated with the generator $\left(\epsilon^{-1}\mathcal{L}_O + \mathcal{L}_H\right)$ is assumed to be

ergodic,[3] the null space of $\left(\epsilon^{-1}\mathcal{L}_O + \mathcal{L}_H\right)$ only contains constant functions, thus (4.52) is true iff

$$\int_{\Omega_{\boldsymbol{x}}} \mathcal{L}_2 \rho_{\boldsymbol{Q},\beta}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = 0. \tag{4.53}$$

Indeed, using the explicit expression (4.50) for $\mathcal{L}_2\rho_{\boldsymbol{Q},\beta}$, (4.53) is easily verified to be true.

Although we have shown the existence of a solution of (4.49), there is no obvious closed form solution for f_2 and general potential functions U. Therefore we proceed as follows: We assume that we can expand f_2 in powers of ϵ, i.e.,

$$f_2 = f_{2,0} + \epsilon f_{2,1} + \epsilon^2 f_{2,2} + \dots O(\epsilon^3),$$

and equate powers of ϵ in (4.49), which yields the following sequence of partial differential equations:

$$\mathcal{L}_O^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,0}) = \left[\frac{1}{4}\beta \boldsymbol{p}^2 \Gamma_{1,1} U''(\boldsymbol{q}) - \frac{1}{4}\Gamma_{1,1}U''(\boldsymbol{q}) + \frac{1}{4}\beta s^T \Gamma_{2,1} \boldsymbol{p} U''(\boldsymbol{q})\right] \rho_{\boldsymbol{Q},\beta}, \tag{4.54}$$

$$\mathcal{L}_H^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,0}) + \mathcal{L}_O^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,1}) = \left[\frac{1}{12}\beta \boldsymbol{p}^3 U^{(3)}(\boldsymbol{q}) \rho_{\boldsymbol{Q},\beta} - \frac{1}{4}\beta \boldsymbol{p} U'(\boldsymbol{q}) U''(\boldsymbol{q})\right] \rho_{\boldsymbol{Q},\beta}, \tag{4.55}$$

$$\mathcal{L}_H^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,1}) + \mathcal{L}_O^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,2}) = 0, \tag{4.56}$$

$$\mathcal{L}_H^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,i}) + \mathcal{L}_O^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,i+1}) = 0, \ i \geq 2. \tag{4.57}$$

In what follows we derive the form of the leading order term $f_{2,0}$. A simple calculation shows that

$$\mathcal{L}_O^\dagger(\rho_{\boldsymbol{Q},\beta} g(\boldsymbol{q},\boldsymbol{p}))$$

with

$$g(\boldsymbol{q},\boldsymbol{p}) = -\frac{1}{8}\beta \boldsymbol{p}^2 U''(\boldsymbol{q})$$

coincides with the left hand side of (4.54), thus $f_{2,0}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = g(\boldsymbol{q},\boldsymbol{p})$ is a solution of (4.54). Moreover, this solution is unique up to a term which is constant in $(\boldsymbol{p},\boldsymbol{s})$, i.e.,

$$f_{2,0}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = -\frac{1}{8}\beta \boldsymbol{p}^2 U''(\boldsymbol{q}) + \phi(\boldsymbol{q}). \tag{4.58}$$

This can be seen as follows. Let ν denote the marginal of $\rho_{\boldsymbol{Q},\beta}$ in $\boldsymbol{q}$ and κ the marginal of $\rho_{\boldsymbol{Q},\beta}$ in $(\boldsymbol{p},\boldsymbol{s})$ so that $\rho_{\boldsymbol{Q},\beta}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = \nu(\boldsymbol{q})\kappa(\boldsymbol{p},\boldsymbol{s})$. Let $\hat{g}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s})$ be another solution of (4.54), then

$$0 = \mathcal{L}_O^\dagger((g-\hat{g})\rho_{\boldsymbol{Q},\beta}) \quad = \mathcal{L}_O^\dagger((g-\hat{g})\rho_{\boldsymbol{Q},\beta}) = \mathcal{L}_O^\dagger((g-\hat{g})\kappa)\nu,$$

thus indeed the difference of g and $\hat{g}$ can only be a function in $\boldsymbol{q}$ since the ergodicity of the Markov process assoicated with $\mathcal{L}^\dagger$ implies the ergodicity of the Ornstein-Uhlenbeck process associated with $\mathcal{L}_O^\dagger$, which in particular implies that the null space of $\mathcal{L}_O^\dagger$ consists functions of the generic form $\kappa(\boldsymbol{p},\boldsymbol{s})\phi(\boldsymbol{q})$.

Substituting $f_{2,0}$ in (4.55) by the right hand side of (4.58), we obtain

$$\mathcal{L}_O^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,1}) = \frac{1}{24}\beta \boldsymbol{p}^3 U^{(3)}(\boldsymbol{q}) \rho_{\boldsymbol{Q},\beta} + \boldsymbol{p} v'(\boldsymbol{q}) \rho_{\boldsymbol{Q},\beta}. \tag{4.59}$$

[3] If the SDE associated with $\mathcal{L}_O + \mathcal{L}_H$ is ergodic it follows directly that also $\left(\epsilon^{-1}\mathcal{L}_O + \mathcal{L}_H\right)$ is ergodic for any finite $\epsilon > 0$.

Again, by the same arguments as used to derive the generic form of $f_{2,0}$, we find that the solution $f_{2,1}$ of (4.59) is uniquely determined up to a function $\Psi(\boldsymbol{q})$, i.e.,

$$f_{2,1}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = g_{\boldsymbol{\Gamma}}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) + \Psi(\boldsymbol{q}), \tag{4.60}$$

where the exact form of the function $g_{\boldsymbol{\Gamma}}$ is provided in appendix C.

By virtue of the Fredholm alternative, there exists a solution for (4.56) if and only if

$$\int_{\Omega_{\boldsymbol{p}}\times\Omega_{\boldsymbol{s}}} \mathcal{L}_H^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,1})\mathrm{d}\boldsymbol{p}\mathrm{d}\boldsymbol{s} = 0.$$

Using the generic form (4.60) of $f_{2,1}$ we find

$$\begin{aligned}\int_{\Omega_{\boldsymbol{p}}\times\Omega_{\boldsymbol{s}}} \mathcal{L}_H^\dagger(\rho_{\boldsymbol{Q},\beta} f_{2,1})\mathrm{d}\boldsymbol{p}\mathrm{d}\boldsymbol{s} = \frac{\Gamma_{2,2}}{\Gamma_{1,1}\Gamma_{2,2}-\Gamma_{1,2}\Gamma_{2,1}}\Big(&\frac{U^{(4)}(\boldsymbol{q})}{8\beta} + U'(\boldsymbol{q})\phi'(\boldsymbol{q}) \\ &- \frac{1}{8}U^{(3)}(\boldsymbol{q})U'(\boldsymbol{q}) - \frac{\phi''(\boldsymbol{q})}{\beta}\Big).\end{aligned} \tag{4.61}$$

Since the right hand side of (4.61) vanishes if and only if we find $\phi(\boldsymbol{q}) = \frac{1}{8}U''(\boldsymbol{q})$, thus

$$f_{2,0}(\boldsymbol{q},\boldsymbol{p},\boldsymbol{s}) = -\frac{1}{8}\beta\boldsymbol{p}^2U''(\boldsymbol{q}) + \frac{1}{8}U''(\boldsymbol{q}).$$

Let φ be a suitable observable which is constant in $(\boldsymbol{p},\boldsymbol{s})$. Since

$$\int_{\Omega_{\boldsymbol{p}}\times\Omega_{\boldsymbol{s}}} f_{2,0}\rho_{\boldsymbol{Q},\beta}\mathrm{d}\boldsymbol{p}\mathrm{d}\boldsymbol{s} = 0,$$

if follows that

$$\int_{\Omega_{\boldsymbol{x}}} \varphi(\boldsymbol{q}) f_{2,0}\rho_{\boldsymbol{Q},\beta}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 0.$$

We conclude

$$\begin{aligned}\int_{\Omega_{\boldsymbol{x}}} \varphi\hat{\rho}_{\Delta t}\mathrm{d}\boldsymbol{x} &= \int_{\Omega_{\boldsymbol{x}}} \varphi\rho_{\boldsymbol{Q},\beta}\mathrm{d}\boldsymbol{x} + \Delta t^2\int_{\Omega_{\boldsymbol{x}}} \varphi f_{2,0}\rho_{\boldsymbol{Q},\beta}\mathrm{d}\boldsymbol{x} \\ &\quad + \epsilon\Delta t^2\int_{\Omega_{\boldsymbol{x}}} \varphi f_{2,1}\rho_{\boldsymbol{Q},\beta}\mathrm{d}\boldsymbol{x} + O(\epsilon^2\Delta t^2) + O(\Delta t^4) \\ &= \int_{\Omega_{\boldsymbol{x}}} \varphi\rho_{\boldsymbol{Q},\beta}\mathrm{d}\boldsymbol{x} + \epsilon\Delta t^2\int_{\Omega_{\boldsymbol{x}}} \varphi f_{2,1}\rho_{\boldsymbol{Q},\beta}\mathrm{d}\boldsymbol{x} + O(\epsilon^2\Delta t^2) + O(\Delta t^4),\end{aligned}$$

which explains (4.44).

Remark 4.6.1. *We reiterate that the above calculations are not rigorous. Apart from the fact that we do not specify the functional space in which we are working when applying Fredholm's alternative, all expansions we work with are formal. A rigorous derivation of the expansion (4.48) as it was performed in [72] for discretisations of the underdamped Langevin equation (see also [1, 18]) would require a more detailed analysis of the spectral properties of $\mathcal{L}_{\mathrm{GLE}}^{-1}$. We expect that this can be done by generalising results provided in [63] and [64, 118]. We, however, defer this to future work.*

4.7 Numerical methods for GLEs with complex memory kernels

In this section we describe how the O-integration step can be modified in order to avoid the computational challenges which arise when the O-step is solved exactly. The premise under which we construct the corresponding algorithms is, that the matrix $\widetilde{\mathbf{\Gamma}}$ is sparse and also has some additional structure, which we will detail below. Under these assumptions the computational costs for one integration step can be reduced from order $O\left((n+m)^3\right)$ to oder $O\left(n+m\right)$. The general approach we are following here is to replace the exact O-integration step by an approximate solution computed by a numerical integrator, denoted as $\hat{\Phi}^O_{\Delta t}$ with associated evolution operator $\hat{P}_{O,\Delta t}$. A numerical integrator for the GLE can then be obtained by substituting the approximate solution step in one of the splitting schemes described in the following three sections, e.g.

$$\hat{P}^{XY\hat{O}YX}_{\Delta t} = \exp\left(\frac{\Delta t}{2}\mathcal{L}_X\right)\exp\left(\frac{\Delta t}{2}\mathcal{L}_Y\right)\hat{P}_{O,\Delta t}\exp\left(\frac{\Delta t}{2}\mathcal{L}_Y\right)\exp\left(\frac{\Delta t}{2}\mathcal{L}_X\right). \quad (4.62)$$

with $X, Y \in \{A, B\}$. We will refer to numerical schemes constructed in this way as composite schemes. We construct $\hat{\Phi}^O_{\Delta t}$ in such a way that it makes efficient use of the sparse structure of the matrix $\widetilde{\mathbf{\Gamma}}$. We propose three classes of integrators; one class based on a splitting of the form

$$\hat{P}_{O,\Delta t} = \exp(\frac{\Delta t}{2}\mathcal{L}_D)\exp(\Delta t\mathcal{L}_F)\exp(\frac{\Delta t}{2}\mathcal{L}_D), \quad (4.63)$$

where $\mathcal{L}_D = -\widetilde{\mathbf{\Gamma}}\boldsymbol{z}\cdot\nabla_{\boldsymbol{p}}$, and $\mathcal{L}_F = \beta^{-1}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{\Sigma}}^T : \nabla^2_{\boldsymbol{p}}$, and another class based on a splitting of the form

$$\hat{P}_{O,\Delta t} = \exp(\frac{\Delta t}{2}\bar{\mathcal{A}})\exp(\Delta t\bar{\mathcal{S}})\exp(\frac{\Delta t}{2}\bar{\mathcal{A}}), \quad (4.64)$$

where $\bar{\mathcal{A}}+\bar{\mathcal{S}} = \mathcal{L}_O$ and $\bar{\mathcal{A}}$ anti-symmetric, e.g. $\bar{\mathcal{A}} = \mathcal{A}, \bar{\mathcal{S}} = \mathcal{S}$, with $\mathcal{A}$ and $\mathcal{S}$ as defined in Proposition 3.1.3. We refer to the splitting (4.63), which is based on a decomposition into a fluctuation and a dissipation part as FD-splitting and to the splitting (4.64), which is based on a decomposition into anti-symmetric and a remainder part as an AR-splitting. Finally, since the vector field associated with O-integration step is linear, we can derive numerical integrators of arbitrary weak order by matching terms of Ito-Taylor expansion of the discrete scheme exact solution.

The remainder of this section is structured as follows. In the following three subsections, we describe the constructions of numerical integrators based on the splittings (4.63), (4.64) and Taylor expansion of the exact flow map, respectively, in detail. In Section 4.7.4 we discuss the stability and weak order accuracy of the integrators. In Section 4.7.5 we discuss computational aspects of the methods presented here. In particular, we discuss the usage of matrix-free operations. In Section 4.7.6 we discuss the use of multiple time-stepping for the approximate solution of the O-step within the GLE integrators. Finally, we provide several detailed description in Section 4.7.7 of adaption of the methods described in this section to applications such as dissipative particle dynamics (DPD) with memory [78, 79] and GLE models for solids [60, 115].

4.7.1 FD-splittings

The stochastic flow map associated with the splitting (4.63) is obtained in a straightforward way as

$$\hat{\Phi}_{\Delta t}(\boldsymbol{z}, \mathscr{R}) = \exp(-\Delta t \widetilde{\boldsymbol{\Gamma}}) z + \sqrt{\Delta t \beta^{-1}} \exp(-\Delta t/2 \widetilde{\boldsymbol{\Gamma}}) \widetilde{\boldsymbol{\Sigma}} \mathscr{R}. \tag{4.65}$$

As this scheme is constructed as a symmetric splitting, we expect the weak error to be of order $O(\Delta t^2)$. A formal expansion of the evolution operator (4.63) yields the leading error term

$$\begin{aligned}
\mathcal{L}_2^{DFD} \varphi &= \frac{1}{12}[\mathcal{L}_F[\mathcal{L}_F, \mathcal{L}_D]]\varphi - \frac{1}{24}[\mathcal{L}_D[\mathcal{L}_D, \mathcal{L}_F]]\varphi \\
&= \frac{1}{3} \boldsymbol{Q} \widetilde{\boldsymbol{\Gamma}}_{1,2}^T \widetilde{\boldsymbol{\Gamma}}_{1,2} \boldsymbol{Q} : \mathrm{D}_{\boldsymbol{s}}\, \varphi - \frac{2}{3} \widetilde{\boldsymbol{\Gamma}}_{2,2} \boldsymbol{Q} \widetilde{\boldsymbol{\Gamma}}_{2,2} : \mathrm{D}_{\mathrm{s}}\, \varphi \\
&\quad - \nabla_{\boldsymbol{p}} \cdot (\widetilde{\boldsymbol{\Gamma}}_{1,2} \boldsymbol{Q} \widetilde{\boldsymbol{\Gamma}}_{2,2} \nabla_{\boldsymbol{s}}\, \varphi) - \frac{1}{3} \boldsymbol{Q} \widetilde{\boldsymbol{\Gamma}}_{1,2} \boldsymbol{Q} \widetilde{\boldsymbol{\Gamma}}_{1,2}^T : \mathrm{D}_{\boldsymbol{p}}\, \varphi.
\end{aligned}$$

Alternatively, one might also consider a splitting of the form FDF. In this case an expansion of the leading error term yields $\mathcal{L}_2^{FDF} \varphi = -\mathcal{L}_2^{DFD} \varphi$.

Remark 4.7.1. *One might be tempted to construct an approximation of the scheme (4.65) as*

$$\hat{\Phi}_{\Delta t}(\boldsymbol{z}, \mathscr{R}) = \left(\boldsymbol{I}_{n+m} - \frac{\Delta t}{2} \widetilde{\boldsymbol{\Gamma}} \right)^{-1} \left(\boldsymbol{z} + \frac{\Delta t}{2} \widetilde{\boldsymbol{\Gamma}} \boldsymbol{z} + \sqrt{\beta^{-1} \Delta t} \widetilde{\boldsymbol{\Sigma}} \mathscr{R} \right), \tag{4.66}$$

which would avoid the computation of the action of the matrix exponential and instead just incorporate one matrix-vector multiplication and the solution of a linear equation. Indeed, if $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{0}$, the scheme (4.66) reduces to the trapezoidal rule, which is second order. However, the convergence order is not preserved in the presence of noise (neither in the strong nor the weak context) and the resulting method (4.66) is not even consistent.

4.7.2 AR-splittings

Employing a splitting of the form (4.64) is motivated by the following observation. If the memory kernel $\boldsymbol{K}$ is of the form

$$\boldsymbol{K}_{i,j}(\boldsymbol{q}, t) = \sum_k (a_{k,i,j})^2(\boldsymbol{q}) K_k(t), \;\; a_{k,i,j}(\boldsymbol{q}) = a_{k,j,i}(\boldsymbol{q}) \in \mathbb{R},$$

where each component $K_k \in \mathcal{C}^\infty(\Omega_{\boldsymbol{q}} \times \mathbb{R}, \mathbb{R})$, is representable in the extended variable form (3.1-3.3), then the matrix $\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q})$ is of the form of what in the numerical linear algebra literature is typically referred to as an "arrow matrix", i.e. $\widetilde{\boldsymbol{\Gamma}}_{2,2}(\boldsymbol{q})$ is a sparse block diagonal matrix and besides the block entries of $\widetilde{\boldsymbol{\Gamma}}_{2,2}(\boldsymbol{q})$ only $\widetilde{\boldsymbol{\Gamma}}_{1,2}(\boldsymbol{q})$ and $\widetilde{\boldsymbol{\Gamma}}_{2,1}(\boldsymbol{q})$ contain non-zero entries. In this case $\exp(-\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{q}))$ is typically dense, while $\exp(-\widetilde{\boldsymbol{\Gamma}}_{2,2}(\boldsymbol{q}))$ has the same block structure as $\widetilde{\boldsymbol{\Gamma}}_{2,2}(\boldsymbol{q})$. More generally, this applies when the memory kernel $\boldsymbol{K}$ is in a *Fast Integrable Kernal Representation*.

Definition 4.7.1 (Fast Integrable Kernel Representation)**.** *We say that the representation of a memory kernel as $\boldsymbol{K}(\boldsymbol{q}, t) = \widetilde{\boldsymbol{\Gamma}}_{1,2} e^{-t\widetilde{\boldsymbol{\Gamma}}_{2,2}} \boldsymbol{B} \widetilde{\boldsymbol{\Gamma}}_{1,2}^T$ is in a Fast Integrable Kernel*

Representation (FIKR), if $\boldsymbol{B}$ *and* $\widetilde{\boldsymbol{\Gamma}}_{2,2}$ *are block diagonal matrices of the form*

$$\boldsymbol{B} = \mathrm{diag}(B_1, \dots B_r),\ \widetilde{\boldsymbol{\Gamma}}_{2,2} = \mathrm{diag}(\tilde{\Gamma}_1, \dots \tilde{\Gamma}_r),$$

where $B_i \tilde{\Gamma}_i(\boldsymbol{q}) \in \mathbb{R}^{d_i \times d_i}$*, and* B_i *symmetric positive definite with* $\sum_{i=1}^r d_i = m$ *and* $d_i \ll m$*, so that*

$$\tilde{\Gamma}_i(\boldsymbol{q})B_i + B_i\tilde{\Gamma}_i^T(\boldsymbol{q})$$

are positive semi-definite for $i = 1, \dots, r$*.*

Now, if $\bar{\mathcal{A}}$ in (4.64) is chosen as

$$\bar{\mathcal{A}} = -\widetilde{\boldsymbol{\Gamma}}_{1,2}\boldsymbol{s} \cdot \nabla_{\boldsymbol{p}} + \boldsymbol{B}\widetilde{\boldsymbol{\Gamma}}_{1,2}^T\boldsymbol{p} \cdot \nabla_{\boldsymbol{s}} =: -\widetilde{\boldsymbol{\Gamma}}^{\bar{\mathcal{A}}}\boldsymbol{z} \cdot \nabla_{\boldsymbol{z}},$$

then the exact flow map

$$\hat{\Phi}_{\Delta t}^{\bar{\mathcal{S}}}(\boldsymbol{s}, \mathscr{R}) = \boldsymbol{F}_{\Delta t}^{\bar{\mathcal{S}}}\boldsymbol{s} + \boldsymbol{S}_{\Delta t}^{\bar{\mathcal{S}}}\mathscr{R},$$

of the linear SDE associated with the remainder part

$$\bar{\mathcal{S}} = -\widetilde{\boldsymbol{\Gamma}}_{2,2}\boldsymbol{s}\nabla_{\boldsymbol{s}} + \nabla_{\boldsymbol{s}}\left(\widetilde{\boldsymbol{\Sigma}}_{2,2}\widetilde{\boldsymbol{\Sigma}}_{2,2}^T\nabla_{\boldsymbol{s}}\ \cdot\ \right)$$

can be computed efficiently as the block structure of $\widetilde{\boldsymbol{\Gamma}}_{2,2}$ is inherited by $\boldsymbol{F}_{\Delta t}^{\bar{\mathcal{S}}}$ and $\boldsymbol{S}_{\Delta t}^{\bar{\mathcal{S}}}$. As $\exp(-\widetilde{\boldsymbol{\Gamma}}^{\bar{\mathcal{A}}})$ is in general dense an exact solution of the differential equation associated with $\bar{\mathcal{A}}$ is not advisable. We therefore further subdivide $\bar{\mathcal{A}}$ as $\bar{\mathcal{A}} = \bar{\mathcal{A}}_A + \bar{\mathcal{A}}_B$, where

$$\bar{\mathcal{A}}_A = \boldsymbol{B}\widetilde{\boldsymbol{\Gamma}}_{1,2}^T\boldsymbol{p} \cdot \nabla_{\boldsymbol{s}}, \quad \bar{\mathcal{A}}_B = -\widetilde{\boldsymbol{\Gamma}}_{1,2}\boldsymbol{s} \cdot \nabla_{\boldsymbol{p}}, \tag{4.67}$$

and approximate the solution corresponding to the splitting (4.64) as

$$\hat{P}_{O,\Delta t}^{\text{AR-2}} = \exp(\frac{\Delta t}{2}\bar{\mathcal{A}}_B)\exp(\frac{\Delta t}{2}\bar{\mathcal{A}}_A)\exp(\Delta t\bar{\mathcal{S}})\exp(\frac{\Delta t}{2}\bar{\mathcal{A}}_A)\exp(\frac{\Delta t}{2}\bar{\mathcal{A}}_B), \tag{4.68}$$

We refer to the numerical method corresponding to the splitting (4.68) as AR-2. We provide an algorithmic implementation of the corresponding scheme as algorithm 3. Note that AR-2 resembles a BAOAB integration sequence and for $m = n$ and $\widetilde{\boldsymbol{\Gamma}}_{1,2} = \boldsymbol{I}_n$ corresponds exactly to a BAOAB integrator for the underdamped Langevin equation with Hamiltonian $H(\boldsymbol{p}, \boldsymbol{s}) = \frac{1}{2}\boldsymbol{p}^T\boldsymbol{p} + \frac{1}{2}\boldsymbol{s}^T\boldsymbol{Q}\boldsymbol{s}$. Indeed, the reason why we construct AR-2 using this particular integration order is exactly due to the fact that the BAOAB integration sequence leads to discrete solutions of particular high accuracy.

Algorithm 3 AR-2

1: INPUT $\boldsymbol{z}, \widetilde{\boldsymbol{\Gamma}}^{\bar{\mathcal{A}}}, \boldsymbol{F}_{\Delta t}^{\bar{\mathcal{S}}}, \boldsymbol{S}_{\Delta t}^{\bar{\mathcal{S}}}, \Delta t$
2: $\boldsymbol{p} \coloneqq \boldsymbol{p} + \frac{\Delta t}{2}\widetilde{\boldsymbol{\Gamma}}_{1,2}^{\bar{\mathcal{A}}}\boldsymbol{s}$
3: $\boldsymbol{s} \coloneqq \boldsymbol{s} + \frac{\Delta t}{2}\widetilde{\boldsymbol{\Gamma}}_{2,1}^{\bar{\mathcal{A}}}\boldsymbol{p}$
4: $\boldsymbol{s} \coloneqq \boldsymbol{F}_{\Delta t}^{\bar{\mathcal{S}}}\boldsymbol{s} + \boldsymbol{S}_{\Delta t}^{\bar{\mathcal{S}}}\mathscr{R},\ \mathscr{R} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m)$
5: $\boldsymbol{s} \coloneqq \boldsymbol{s} + \frac{\Delta t}{2}\widetilde{\boldsymbol{\Gamma}}_{2,1}^{\bar{\mathcal{A}}}\boldsymbol{p}$
6: $\boldsymbol{p} \coloneqq \boldsymbol{p} + \frac{\Delta t}{2}\widetilde{\boldsymbol{\Gamma}}_{1,2}^{\bar{\mathcal{A}}}\boldsymbol{s}$
7: **return** $\boldsymbol{z}$

4.7.3 Construction by direct moment approximation

Integrators up to arbitrary high orders can be constructed via a weak (Ito-) Taylor expansion [87, 62]. For the stochastic process at hand the construction is drastically simplified as the corresponding the SDE is linear with an additive noise term and hence the law of the solution for a deterministic initial condition, i.e. $\boldsymbol{z}_0 \sim \delta_{\boldsymbol{z}}(\cdot)$, is Gaussian. Provided that the law of the discrete approximation $\hat{\Phi}_{\Delta t}(\boldsymbol{z})$ is Gaussian as well, this means in particular that in order to construct an integrator of order p, it is sufficient to ensure that the first and second moments of $\hat{\Phi}_{\Delta t}(\boldsymbol{z})$ approximate the corresponding moments of the exact solution at order $O(\Delta t^p)$, i.e.

$$\mathbb{E}\left[\hat{\Phi}_{\Delta t}(\boldsymbol{z})\right] = \mathbb{E}\left[\Phi^O_{\Delta t}(\boldsymbol{z})\right] + O(\Delta t^{p+1}), \boldsymbol{z} \in \Omega_{\boldsymbol{z}}, \tag{4.69}$$

and

$$\mathbb{E}\left[\hat{\Phi}_{\Delta t}(\boldsymbol{z})\hat{\Phi}_{\Delta t}(\boldsymbol{z})^T\right] = \mathbb{E}\left[\Phi^O_{\Delta t}(\boldsymbol{z})\Phi^O_{\Delta t}(\boldsymbol{z})^T\right] + O(\Delta t^{p+1}), \boldsymbol{z} \in \Omega_{\boldsymbol{z}}, \tag{4.70}$$

where $\Phi^O_{\Delta t}$ as defined in (4.7). The first and second moment of the exact solution can be expanded in Δt as

$$\mathbb{E}\left[\Phi^O_{\Delta t}(\boldsymbol{z})\right] = \sum_{k=0}^{p} \frac{\Delta t^k}{k!} \partial_t^k e^{-t\widetilde{\boldsymbol{\Gamma}}} z \Big|_{t=0} + O(\Delta t^{p+1}),$$

and

$$\mathbb{E}\left[\Phi^O_{\Delta t}(\boldsymbol{z})\Phi^O_{\Delta t}(\boldsymbol{z})^T\right] = \sum_{k=0}^{p} \Delta t^k \frac{1}{k!} \partial_t^k \left((e^{-t\widetilde{\boldsymbol{\Gamma}}} z)(e^{-t\widetilde{\boldsymbol{\Gamma}}} z)^T\right)\Big|_{t=0} + \sum_{k=0}^{p} \Delta t^k \boldsymbol{A}_k, + O(\Delta t^{p+1}), \tag{4.71}$$

where

$$\begin{aligned} \boldsymbol{A}_k = \frac{1}{k!}\partial_t^k \left(\boldsymbol{Q} - e^{-t\widetilde{\boldsymbol{\Gamma}}}\boldsymbol{Q}e^{-t\widetilde{\boldsymbol{\Gamma}}^T}\right)\Big|_{t=0} &= \frac{1}{k!}\sum_{l=0}^{k}(-1)^k \binom{k}{l} \widetilde{\boldsymbol{\Gamma}}^{k-l}\boldsymbol{Q}\left(\widetilde{\boldsymbol{\Gamma}}^T\right)^l \\ &= \frac{1}{k!}\sum_{l=0}^{k-1}(-1)^{k-1}\widetilde{\boldsymbol{\Gamma}}^{k-1-l}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Sigma}}^T\left(\widetilde{\boldsymbol{\Gamma}}^T\right)^l. \end{aligned} \tag{4.72}$$

Assuming $\hat{\Phi}_{\Delta t}$ to be of the generic form

$$\hat{\Phi}_{\Delta t}(\boldsymbol{z}) = \sum_{k=0}^{p} \Delta t^k \hat{\boldsymbol{f}}_k \boldsymbol{z} + \sum_{k=1}^{p} \Delta t^{\frac{2k-1}{2}} \sum_{j=1}^{n_p} \hat{\boldsymbol{g}}_{k,j} \mathscr{R}_j, \tag{4.73}$$

where $n_p \geq 1$, $\mathscr{R}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{n+m})$ i.i.d., and $\boldsymbol{f}_k, \boldsymbol{g}_k \in \mathbb{R}^{(n+m)\times(n+m)}$, we find

$$\mathbb{E}\left[\hat{\Phi}_{\Delta t}(\boldsymbol{z})\right] = \sum_{k=0}^{p} \Delta t^k \boldsymbol{f}_k \boldsymbol{z}, \tag{4.74}$$

and

$$\mathbb{E}\left[\hat{\Phi}_{\Delta t}(\boldsymbol{z})\hat{\Phi}_{\Delta t}(\boldsymbol{z})^T\right] = \sum_{k=0}^{2p} \Delta t^k \sum_{l=0}^{k} (\boldsymbol{f}_k \boldsymbol{z})(\boldsymbol{f}_{k-l}\boldsymbol{z})^T + \sum_{k=1}^{2p} \Delta t^k \sum_{l=1}^{k}\sum_{j=1}^{n_p} \boldsymbol{g}_{l,j}\boldsymbol{g}_{k-l,j}^T. \tag{4.75}$$

p	$\boldsymbol{g}_{1,1}$	$\boldsymbol{g}_{2,1}$	$\boldsymbol{g}_{3,1}$	$\boldsymbol{g}_{4,1}$	$\boldsymbol{g}_{2,2}$	$\boldsymbol{g}_{3,2}$
1	$\widetilde{\boldsymbol{\Sigma}}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$
2	$\widetilde{\boldsymbol{\Sigma}}$	$-\frac{1}{2}\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Sigma}}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$
3	$\widetilde{\boldsymbol{\Sigma}}$	$-\frac{1}{2}\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Sigma}}$	$\frac{1}{6}\widetilde{\boldsymbol{\Gamma}}^2\widetilde{\boldsymbol{\Sigma}}$	$\mathbf{0}$	$\frac{1}{\sqrt{12}}\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Sigma}}$	$\mathbf{0}$
4	$\widetilde{\boldsymbol{\Sigma}}$	$-\frac{1}{2}\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Sigma}}$	$\frac{1}{6}\widetilde{\boldsymbol{\Gamma}}^2\widetilde{\boldsymbol{\Sigma}}$	$-\frac{1}{24}\widetilde{\boldsymbol{\Gamma}}^3\widetilde{\boldsymbol{\Sigma}}$	$\frac{1}{\sqrt{12}}\widetilde{\boldsymbol{\Gamma}}\widetilde{\boldsymbol{\Sigma}}$	$-\frac{1}{\sqrt{12}}\widetilde{\boldsymbol{\Gamma}}^2\widetilde{\boldsymbol{\Sigma}}$

Table 4.1: Parametrization of the random terms in (4.73) for schemes up to weak order $p = 4$.

Now, by equating powers of Δt values of the coefficients $\boldsymbol{g}_{k,j}$ and $\boldsymbol{f}_k$ can be derived. Comparing (4.69) and (4.74) one directly finds

$$\boldsymbol{f}_k = \frac{(-1)^k}{k!}\widetilde{\boldsymbol{\Gamma}}^k,$$

for $k \leq p$, which also leads to an error of order Δt^{2p} in the z-dependent term in the second moment. The values of the coefficients $\boldsymbol{g}_{l,j}$, must be determined by direct inspection. We provide the values for the construction of schemes up to 4th order in table Table 4.1. To our knowledge there does not exists a simple closed form solution for these coefficients.

Remark 4.7.2. *If the matrices $\widetilde{\boldsymbol{\Gamma}}$ and $\widetilde{\boldsymbol{\Sigma}}$ commute, one finds that only a single random vector, $\mathscr{R}_1$ (i.e. $\boldsymbol{g}_{l,j} = 0$ for $j > 1$) is required to construct a numerical scheme of arbitrary high order p using the above described approach. Indeed, let $p \geq 1$; by equating powers of Δt in the expressions of (4.71) and (4.75) so that terms up order $p+1$ coincide, one can determine the values of the coefficients $\boldsymbol{g}_{k,j}$ so that the choice*

$$\boldsymbol{g}_{k,j} = \begin{cases} a_k(\widetilde{\boldsymbol{\Gamma}} + \widetilde{\boldsymbol{\Gamma}}^T)^{k-1}\widetilde{\boldsymbol{\Sigma}} & \text{if } k \leq p \text{ and } j = 1, \\ \mathbf{0} & \text{otherwise} \end{cases}$$

with $a_1 = 1, a_2 = -\frac{1}{2}, a_3 = \frac{5}{95}, a_4 = \frac{1}{128}, a_5 = \frac{79}{92160}, a_6 = \frac{3}{40960}$, results in a scheme of at least weak order p.

4.7.4 Stability and weak order accuracy

In the remainder of this section let $\hat{\boldsymbol{F}}_{\Delta t}$ denote the method dependent matrix for which $\boldsymbol{z} \mapsto \mathbb{E}\left[\hat{\Phi}^O_{\Delta t}\boldsymbol{z}\right] = \hat{\boldsymbol{F}}_{\Delta t}\boldsymbol{z}$ holds. The ergodicity of H-OU schemes relies on the existence of a suitable Lyapunov function. We showed in Section 4.4 for the case $\Omega_{\boldsymbol{q}} = L\mathbb{T}^n$, that if an exact solution of the O-part is used, then $\mathcal{K}_l(\boldsymbol{x}) = (\boldsymbol{x}^T\boldsymbol{C}\boldsymbol{x})^l, l \in \mathbb{N}$, where $\boldsymbol{C}$ is a positive definite matrix such that $\widetilde{\boldsymbol{\Gamma}}\boldsymbol{C} + \boldsymbol{C}\widetilde{\boldsymbol{\Gamma}}^T$ is positive definite, satisfies the uniform Lyapunov condition (Assumption 9) for sufficiently small step size Δt. The proof of this results relies on $\boldsymbol{z} \mapsto \mathbb{E}\left[\Phi^O_{\Delta t}\boldsymbol{z}\right] = \boldsymbol{F}_{\Delta t\Delta t}\boldsymbol{z}$ to define a contraction in $(\Omega_{\boldsymbol{z}}, \|\cdot\|_{\boldsymbol{C}})$. Now, if instead an approximate solution of the O-part is used, the corresponding map $\boldsymbol{z} \mapsto \mathbb{E}\left[\hat{\Phi}^O_{\Delta t}\boldsymbol{z}\right] = \hat{\boldsymbol{F}}_{\Delta t}\boldsymbol{z}$ does not in general define a contraction.

Lemma 4.7.1. *Under the same assumptions as Proposition 4.4.1, let $\Delta t \leq \Delta t^*$, where Δt^* as defined in Proposition 4.4.1. If there exists a symmetric positive definite matrix $\boldsymbol{C}_{\Delta t}$, such that $\boldsymbol{z} \mapsto \hat{\boldsymbol{F}}_{\Delta t}\boldsymbol{z}$ defines a contraction in $\left(\Omega_{\boldsymbol{z}}, \|\cdot\|_{\boldsymbol{C}_{\Delta t}}\right)$, then the scheme associated with the evolution operator (4.62) is ergodic.*

Proof. Following the same steps as in the proof of Proposition 4.4.1, the contractivity of $\boldsymbol{z} \mapsto \hat{\boldsymbol{F}}_{\Delta t}\boldsymbol{x}$ allows to show that $\mathcal{K}_l(\boldsymbol{z}) = (\boldsymbol{z}^T\boldsymbol{C}_{\Delta t}\boldsymbol{z})^l$ defines a family of Lyapunov functions so that Assumption 9 is satisfied, which implies the existence of a limiting probability measure with positive density. □

We recall that all eigenvalues of a matrix $A \in \mathbb{R}^m$ lie within the complex unit circle if and only if the discrete Lyapunov equation

$$A^T\boldsymbol{X}A - \boldsymbol{X} = -\boldsymbol{I}_m$$

has positive definite matrix $\boldsymbol{X}$ as a solution (see e.g. [129]), hence a matrix $\boldsymbol{C}_{\Delta t}$ with the properties as required in Lemma 4.7.1 exists if and only if the moduli of all eigenvalues of $\hat{\boldsymbol{F}}_{\Delta t}$ are less than 1. For FD-2 and ST-p-exp, one finds $\mathbb{E}\left[\hat{\Phi}^O_{\Delta t}(\boldsymbol{z})\right] = \exp(-\widetilde{\boldsymbol{\Gamma}}\Delta t)\boldsymbol{z} = \boldsymbol{F}_{\Delta t}\boldsymbol{z}$. It therefore follows for both methods that the composite scheme is stable under the same conditions as in the case of an exact solution of the O-step. For ST-p, we have $\hat{\boldsymbol{F}}_{\Delta t} = \sum_{k=0}^{p} \frac{(-1)^k\Delta t^k}{k!}\widetilde{\boldsymbol{\Gamma}}^k$, which means that stability of the composite scheme can only be expected if

$$\sigma_{\max}\left(\sum_{k=0}^{p} \frac{(-1)^k\Delta t^k}{k!}\widetilde{\boldsymbol{\Gamma}}^k\right) < 1.$$

Finally, for AR-2 the analysis is slightly more involved.

Proposition 4.7.1. *For AR-2, there exists a symmetric positive matrix $\boldsymbol{C}_{\Delta t}$ so that, $\boldsymbol{z} \mapsto \mathbb{E}\left[\hat{\Phi}^{AR\text{-}2}_{O,\Delta t}(\boldsymbol{z})\right] = \hat{\boldsymbol{F}}_{\Delta t}z$ is a contraction in $\left(\Omega_{\boldsymbol{z}}, \|\cdot\|_{\boldsymbol{C}_{\Delta t}}\right)$, if*

$$\Delta t < \sigma_{\max}(\widetilde{\boldsymbol{\Gamma}}_{1,2}\boldsymbol{Q}\widetilde{\boldsymbol{\Gamma}}^T_{1,2})^{-1/2}.$$

Proof. Let $\Phi^X_{\Delta t}$ denote the (stochastic) flow map associated with the evolution operators $\exp(-Xt)$ with $X \in \{\bar{\mathcal{A}}_A, \bar{\mathcal{A}}_B, \bar{\mathcal{S}}\}$, respectively and denote by $\boldsymbol{\Psi}^X_{\Delta t} \in \mathbb{R}^{(n+m)\times(n+m)}$ the corresponding matrix such that

$$\boldsymbol{\Psi}^X_{\Delta t}z = \mathbb{E}\left[\Phi^X_{\Delta t}(\boldsymbol{z})\right].$$

For the scheme AR-2 we can write $\hat{\boldsymbol{F}}_{\Delta t}$ as

$$\begin{aligned}\hat{\boldsymbol{F}}_{\Delta t} &= \boldsymbol{\Psi}^{\bar{\mathcal{A}}_B}_{\Delta t/2}\boldsymbol{\Psi}^{\bar{\mathcal{A}}_A}_{\Delta t/2}\boldsymbol{\Psi}^{\bar{\mathcal{S}}}_{\Delta t}\boldsymbol{\Psi}^{\bar{\mathcal{A}}_A}_{\Delta t/2}\boldsymbol{\Psi}^{\bar{\mathcal{A}}_B}_{\Delta t/2} \\ &= \left(\boldsymbol{\Psi}^{\bar{\mathcal{A}}_B}_{\Delta t/2}\boldsymbol{\Psi}^{\bar{\mathcal{A}}_A}_{\Delta t/2}\boldsymbol{\Psi}^{\bar{\mathcal{S}}}_{\Delta t/2}\right)\boldsymbol{\Psi}^{\bar{\mathcal{S}}}_{\Delta t/2}\boldsymbol{A}_{\Delta t}\boldsymbol{\Psi}^{\bar{\mathcal{S}}}_{\Delta t/2}\left(\boldsymbol{\Psi}^{\bar{\mathcal{A}}_B}_{\Delta t/2}\boldsymbol{\Psi}^{\bar{\mathcal{A}}_A}_{\Delta t/2}\boldsymbol{\Psi}^{\bar{\mathcal{S}}}_{\Delta t/2}\right)^{-1},\end{aligned} \tag{4.76}$$

with

$$\boldsymbol{A}_{\Delta t} = \boldsymbol{\Psi}^{\bar{\mathcal{A}}_A}_{\Delta t/2}\boldsymbol{\Psi}^{\bar{\mathcal{A}}_B}_{\Delta t}\boldsymbol{\Psi}^{\bar{\mathcal{A}}_A}_{\Delta t/2},$$

hence the result follows if we can show that $\boldsymbol{z} \mapsto \boldsymbol{\Psi}^{\bar{\mathcal{S}}}_{\Delta t/2}\boldsymbol{A}_{\Delta t}\boldsymbol{\Psi}^{\bar{\mathcal{S}}}_{\Delta t/2}\boldsymbol{z}$ defines a contraction in a suitable normed space. The matrix $\boldsymbol{A}_{\Delta t}$ can be computed as

$$\boldsymbol{A}_{\Delta t} = \begin{pmatrix} \boldsymbol{I}_n - \frac{1}{2}\Delta t^2\boldsymbol{G} & \Delta t\left(\widetilde{\boldsymbol{\Gamma}}_{1,2} - \frac{\Delta t^2}{4}\boldsymbol{G}\widetilde{\boldsymbol{\Gamma}}_{1,2}\right) \\ -\Delta t\boldsymbol{Q}\widetilde{\boldsymbol{\Gamma}}^T_{1,2} & \boldsymbol{I}_m - \frac{\Delta t^2}{2}\boldsymbol{Q}\widetilde{\boldsymbol{\Gamma}}^T_{1,2}\widetilde{\boldsymbol{\Gamma}}_{1,2} \end{pmatrix},$$

where $\boldsymbol{G} \coloneqq \widetilde{\boldsymbol{\Gamma}}_{1,2}\boldsymbol{Q}\widetilde{\boldsymbol{\Gamma}}_{1,2}^T$, and we find that for

$$\boldsymbol{C}_{\Delta t} \coloneqq \begin{pmatrix} \left(\boldsymbol{I}_n - \frac{\Delta t^2}{4}\boldsymbol{G}\right)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}^{-1} \end{pmatrix}.$$

one has

$$\boldsymbol{A}_{\Delta t}^T \boldsymbol{C}_{\Delta t} \boldsymbol{A}_{\Delta t} - \boldsymbol{C}_{\Delta t} = 0.$$

For $\Delta t < \sigma_{\max}(\widetilde{\boldsymbol{\Gamma}}_{1,2}\boldsymbol{Q}\widetilde{\boldsymbol{\Gamma}}_{1,2}^T)^{-1/2}$, the matrix $\boldsymbol{C}_{\Delta t}$ is positive definite and for the corresponding norm $\|\boldsymbol{A}_{\Delta t} z\|_{\boldsymbol{C}_{\Delta t}} = \|\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}}$ holds. Also, for $\boldsymbol{z} = (\boldsymbol{p}, \boldsymbol{s})$ we have

$$\|\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}}^2 = \boldsymbol{p}\left(\boldsymbol{I}_n - \frac{\Delta t^2}{4}\boldsymbol{G}\right)^{-1}\boldsymbol{p}^T + \boldsymbol{s}^T \boldsymbol{F}_{\Delta t}{}^{\bar{\mathcal{S}}^T} \boldsymbol{Q}^{-1} \boldsymbol{F}_{\Delta t}{}^{\bar{\mathcal{S}}} \boldsymbol{s},$$

so

$$\|\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}} = \|\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}} \iff \boldsymbol{s} = 0, \tag{4.77}$$

and

$$\|\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}} < \|\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}}, \tag{4.78}$$

otherwise. Now, assume there is $\boldsymbol{z} = (\boldsymbol{p}, \boldsymbol{s})$ so that $\|\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}} \leq \|\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{A}_{\Delta t}\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}} z\|_{\boldsymbol{C}_{\Delta t}}$. Using (4.77) and (4.78) we find

$$\|\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}} \leq \|\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{A}_{\Delta t}\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}} \leq \|\boldsymbol{A}_{\Delta t}\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}} = \|\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}},$$

hence $\boldsymbol{s} = 0$. Similarly,

$$\|\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}} \leq \|\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{A}_{\Delta t}\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}} \leq \|\boldsymbol{\Psi}_{\Delta t/2}^{\bar{\mathcal{S}}}\boldsymbol{A}_{\Delta t}\boldsymbol{z}\|_{\boldsymbol{C}_{\Delta t}},$$

implies $-\Delta t \boldsymbol{Q}\widetilde{\boldsymbol{\Gamma}}_{1,2}^T\boldsymbol{p} + (\boldsymbol{I}_m - \frac{\Delta t^2}{2}\boldsymbol{Q}\widetilde{\boldsymbol{\Gamma}}_{1,2}^T\widetilde{\boldsymbol{\Gamma}}_{1,2})\boldsymbol{s} = 0$. Together we conclude $\boldsymbol{z} = 0$. □

Remark 4.7.3. *Note that no statement about the uniqueness of the limiting distribution is made in Lemma 4.7.1. We expect however that a minorization condition and thereby the irreducibility of the corresponding Markov chain can be derived analogously as in Section 4.4. However, unlike in the proof of the Lyapunov condition, this would require some additional non-trivial steps. Assuming uniqueness of the invariant distribution, exponential convergence rates might depend on the value of Δt in the case of AR-2 and TS-p as in both these cases the form of $\boldsymbol{C}_{\Delta t}$ and therefore the Lyapunov function $\mathcal{K}_l$ used in the proof of Lemma 4.7.1 depend on the value of Δt.*

4.7.5 Computational aspects

The performance of the numerical integrators introduced in Section 4.7.1-Section 4.7.2 can vary dramatically depending on the properties of the memory kernel $\boldsymbol{K}$ and its Markovian representation. In order to be able to efficiently make use of the sub-splitting strategies described in this section it is therefore important to understand what operations are involved in each of the numerical integrators and how the computational costs associated with each of these operations are affected by different structural properties of $\widetilde{\boldsymbol{\Gamma}}$. Similarly, is also important to be able to make predictions about the performance of the subsplitting strategies in comparison to schemes involving an exact solution of the O-step. In this section we provide guidelines for the choice of the numerical method by breaking down the expected costs for one integration step of each of these methods

	general sparse					FIKR	quasi-diagonal
Operation	Exact	ST-2	ST-2-exp	AR-2	DF-2	AR-2	Exact
Matrix exponential (explicit)	(1)	0	0	(1)*	0	(1)*	(1)
Dense matrix-matrix multiplication	(2)	0	0	(2)*	0	0	0
Dense matrix-vector multiplication	2	0	0	2	0	0	0
Dense Cholesky decomposition	(1)	0	0	0	0	0	0
Matrix exponential (matrix-free)	0	0	1	0	2	0	0
Sparse matrix-matrix multiplication	0	0	0	0	0	(2)*	(2)
Sparse matrix-vector multiplication	0	4	2	0	1	2	2
Sparse Cholesky decomposition	0	(1)	(1)	(1)*	(1)	(1)*	(1)

Table 4.2: Matrix operations required for the computation of the exact solution of the O-step and the 2nd order algorithms described in this section. Brackets indicate that the respective operations are only required to be executed during runtime if $\widetilde{\mathbf{\Gamma}}$ is a function of $\boldsymbol{q}$. An additional asterisk symbol indicates that the operation is only required to be executed during runtime if $\widetilde{\mathbf{\Gamma}}_{2,2}$ is a function of $\boldsymbol{q}$. Note that the matrix exponential in AR-2 for FIKR is sparse.

in three different setups,

1. the matrices $\widetilde{\mathbf{\Gamma}}$ and $\widetilde{\mathbf{\Sigma}}$ are sparse with no additional structure assumed. We refer to this type of matrices as *general sparse* matrices. Table 4.2),
2. the matrix $\widetilde{\mathbf{\Gamma}}$ is in a *fast integrable kernel representation* (see Section 4.7.2),
3. the number of disconnected subgraphs of the associated connectivity graph of $\widetilde{\mathbf{\Gamma}}$ is less than or approximately of the same magnitude as $n+m = \dim(\Omega_{\boldsymbol{z}})$. In this case we refer to $\widetilde{\mathbf{\Gamma}}$ as *quasi diagonal*.

The main (computational intensive) operations involved in the computation of either an exact solution of the O-step or an approximation using one of the algorithms introduced in Section 4.7.1-Section 4.7.2 are listed in Table 4.2.

4.7.6 Multiple time-stepping methods

When AR-2 or ST-2 integrators are used as an appproximation of the O-step it can happen that the maximum admissible stable stepsize for $\hat{\Phi}_{O,\Delta t}$ is smaller than the maximum stable stepsize, which can be used in corresponding splitting scheme for the GLE when an exact solution for the O-step is used. Similarly, using an approximation of the O-step might lead to a substantial increase in the discretisation error in comparison to the discretisation error incurred in the numerical scheme when an exact solution of the O-step is used. By using a multiple time-stepping approach we can both avoid restricting maximum step size and on the same time increase the accuracy of the approximation.

Algorithm 4 gle-BA$\hat{\text{O}}^l$AB

INPUT: $\boldsymbol{x}$
$\boldsymbol{p} \coloneqq \boldsymbol{p} - \frac{\Delta t}{2}\nabla_{\boldsymbol{q}} U(\boldsymbol{q})$
$\boldsymbol{q} \coloneqq \boldsymbol{q} + \frac{\Delta t}{2}\boldsymbol{q}$
for $i = 0$ **to** $l - 1$ **do**
  $\boldsymbol{z} \coloneqq \hat{\Phi}_{O,\Delta t/l}(\boldsymbol{z})$
end for
$\boldsymbol{q} \coloneqq \boldsymbol{q} + \frac{\Delta t}{2}\boldsymbol{q}$
$\boldsymbol{p} \coloneqq \boldsymbol{p} - \frac{\Delta t}{2}\nabla_{\boldsymbol{q}} U(\boldsymbol{q})$
return $\boldsymbol{x}$

4.7.7 Application in DPD and modelling of solids

In this section we present two classes of memory kernels, whose structures are particular amenable for an efficient integration via the AR-2 scheme. In Example 4.7.1 we describe the structure of memory kernels in DPD simulations. Memory kernels of the form as presented in Example 4.7.2 arise in the modelling of heat baths (See [60, 115, 93, 94].)

Example 4.7.1. *In recently proposed dissipative particle dynamics (DPD) models, which incorporate memory effects (See [79] also [78]), the memory kernel* $\boldsymbol{K}$ *has the particular form*

$$\boldsymbol{K}(\boldsymbol{q}, t) = \sum_{I \neq J} K(t)\omega^D(\boldsymbol{q}_{I,J})\hat{\boldsymbol{q}}_{I,J} \cdot \hat{\boldsymbol{q}}_{I,J}^T,$$

where

$$\hat{\boldsymbol{q}}_{I,J} \coloneqq \boldsymbol{q}_{I,J}||\boldsymbol{q}_{I,J}||^{-1},$$

with

$$\boldsymbol{q}_{I,J} \coloneqq \boldsymbol{q}_J - \boldsymbol{q}_I.$$

The function $\omega^D \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R})$ *is a non-negative continuously differentiable weight function with compact support modelling the strength of random force interactions between particles. The function*

$$K(t) = -\Gamma_{1,2}e^{-t\Gamma_{2,2}}\Gamma_{2,1} \in \mathbb{R},$$

with suitable matrices $\Gamma_{1,2}, \Gamma_{2,2}, \Gamma_{2,1}$*, models memory effects of the random force between particles pairs. In the above notation* $I, J \in \mathcal{I}$ *denote particle indices. That is* $\mathcal{I} \subset \mathbb{N}^3$*, and* $\boldsymbol{q}_I \coloneqq \Pi_I \boldsymbol{q}$*, where* Π_I *denotes the orthogonal projection operator on the subspace spanned by the canonical basis vectors* $\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_2}, \boldsymbol{e}_{i_3}$ *defined for* $I = (i_1, i_2, i_3)$*. Below we provide an algorithmic implementation of the DPD version of the AR-2 scheme. The input* A *denotes the set of all interacting particle pairs, i.e.,* $A = \{(I, J) : |\boldsymbol{q}_{I,J}| < r\}$*, where* r *is a cutoff radius associated with the weight function* ω^D*.*

Example 4.7.2. *Let*

$$\widetilde{\boldsymbol{\Gamma}} = \begin{pmatrix} \boldsymbol{0} & \widetilde{\boldsymbol{\Gamma}}_{1,2} \\ -\boldsymbol{D}\widetilde{\boldsymbol{\Gamma}}_{1,2}^T & \widetilde{\boldsymbol{\Gamma}}_{2,2} \end{pmatrix}, \text{ and } \widetilde{\boldsymbol{\Sigma}} = \sqrt{2}\begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_m \end{pmatrix},$$

where

$$\widetilde{\boldsymbol{\Gamma}}_{2,2} = \operatorname{diag}(A_1, \dots A_{m/2}), \text{ with } A_i = \begin{pmatrix} \lambda_i & -\alpha_i \\ \alpha_i & \lambda_i \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

and $\lambda_i > 0$ *and* α_i *as well as the coefficients* $\widetilde{\boldsymbol{\Gamma}}_{1,2}$ *are functions of* $\boldsymbol{q}$*. It can be easily*

Algorithm 5 AR-2 (DPD version)

INPUT $\boldsymbol{z}, \Gamma^{\mathcal{A}}, \Gamma^{\mathcal{S}}, F^{\mathcal{S}}_{\Delta t}, S^{\mathcal{S}}_{\Delta t}, \Delta t, A$
for all $(I,J) \in A$ **do**
  $\boldsymbol{s}_{IJ} \coloneqq \boldsymbol{s}_{IJ} + \frac{\Delta t}{2}\Gamma^{\mathcal{A}}_{2,1}\omega^D(\boldsymbol{q}_{I,J})^{\frac{1}{2}}\hat{\boldsymbol{q}}_{I,J}\hat{\boldsymbol{p}}_{I,J}$
end for
for all $(I,J) \in A$ **do**
  $\boldsymbol{p}_I \coloneqq \boldsymbol{p}_I + \frac{\Delta t}{2}\omega^D(\boldsymbol{q}_{I,J})^{\frac{1}{2}}\hat{\boldsymbol{q}}_{I,J}\Gamma^{\mathcal{A}}_{1,2}\hat{\boldsymbol{s}}_{IJ}$
  $\boldsymbol{p}_J \coloneqq \boldsymbol{p}_J - \frac{\Delta t}{2}\omega^D(\boldsymbol{q}_{I,J})^{\frac{1}{2}}\hat{\boldsymbol{q}}_{I,J}\Gamma^{\mathcal{A}}_{1,2}\hat{\boldsymbol{s}}_{IJ}$
end for
for all $(I,J) \in A$ **do**
  $\boldsymbol{s}_{IJ} \coloneqq F_{\Delta t}\boldsymbol{s}_{IJ} + S_{\Delta t}\mathscr{R}^{IJ}, \quad \mathscr{R}^{IJ} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$
end for
for all $(I,J) \in A$ **do**
  $\boldsymbol{p}_I \coloneqq \boldsymbol{p}_I + \frac{\Delta t}{2}\omega^D(\boldsymbol{q}_{I,J})^{\frac{1}{2}}\hat{\boldsymbol{q}}_{I,J}\Gamma^{\mathcal{A}}_{1,2}\hat{\boldsymbol{s}}_{IJ}$
  $\boldsymbol{p}_J \coloneqq \boldsymbol{p}_J - \frac{\Delta t}{2}\omega^D(\boldsymbol{q}_{I,J})^{\frac{1}{2}}\hat{\boldsymbol{q}}_{I,J}\Gamma^{\mathcal{A}}_{1,2}\hat{\boldsymbol{s}}_{IJ}$
end for
for all $(I,J) \in A$ **do**
  $\boldsymbol{s}_{IJ} \coloneqq \boldsymbol{s}_{IJ} + \frac{\Delta t}{2}\Gamma^{\mathcal{A}}_{2,1}\omega^D(\boldsymbol{q}_{I,J})^{\frac{1}{2}}\hat{\boldsymbol{q}}_{I,J}\hat{\boldsymbol{p}}_{I,J}$
end for
return $\boldsymbol{z}$

checked that condition (3.5) is satisfied for $\tilde{\boldsymbol{Q}} = \operatorname{diag}\left(\boldsymbol{M}^{-1}, \boldsymbol{D}^{-1}\right)$. *A very effective composition of* $\widetilde{\boldsymbol{\Gamma}}$ *in terms of the AR-splitting method is given by*

$$\widetilde{\boldsymbol{\Gamma}}^{\bar{\mathcal{S}}} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{\Gamma}}_{2,2} \end{pmatrix}, \ \widetilde{\boldsymbol{\Gamma}}^{\bar{\mathcal{A}}} = \begin{pmatrix} \boldsymbol{0} & \widetilde{\boldsymbol{\Gamma}}_{1,2} \\ -\widetilde{\boldsymbol{\Gamma}}^T_{1,2} & \boldsymbol{0} \end{pmatrix}.$$

as the matrix exponential of each block in $\widetilde{\boldsymbol{\Gamma}}_{2,2}$ *has the simple form*

$$\exp(-A_i t) = \begin{pmatrix} e^{-t\lambda_i}\cos(t\alpha_i) & e^{-t\lambda_i}\sin(t\alpha_i) \\ -e^{-t\lambda_i}\sin(t\alpha_i) & e^{-t\lambda_i}\cos(t\alpha_i) \end{pmatrix},$$

which requires just the evaluation of one scalar exponential and two trigonometric functions.

4.8 Numerical experiments

In this section we access the performance of the splitting methods which we introduced in Section 4.1 in numerical experiments. We only consider instances of the Markovian reformulation (3.1-3.3) of the GLE and we choose the matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ such that for a suitable potential functions U the SDE (3.1-3.3) is ergodic with invariant measure $\mu(\mathrm{d}\boldsymbol{x}) := \mu_{\boldsymbol{Q},\beta}(\mathrm{d}\boldsymbol{x}) = \rho_{\boldsymbol{Q},\beta}(\boldsymbol{x})\mathrm{d}x$ with $\rho_{\boldsymbol{Q},\beta}$ as specified in Proposition 3.1.1. We consider two different choices of the potential function U.

(i) A simple double-well potential defined on $\mathbb{R}$,

$$U_{DW}(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{q}^2 + \sin(1/4 + 2\boldsymbol{q}), \tag{4.79}$$

(ii) The potential function U_{GM} corresponding to the posterior density of a Bayesian Gaussian mixture model similar to the one described in Section 1.2.1. More specifically, we consider the application of a Bayesian Gaussian mixture model to the Hidalgo stamp dataset [52], which consists of the measurements of the thickness of 482 postage stamps from the 1872 Hidalgo issue of post stamps. We parametrise the model as described in [24]; (See also [103] and [56]). That is, we choose the number of components as $N_c = 3$ and assume the posterior distribution to be of the form

$$p(\boldsymbol{\theta} \mid (y_i, z_i)_{1\leq i\leq N}) = p_{\mathrm{prior}}(\boldsymbol{\theta}) p\left(\left(\boldsymbol{w}_k, \boldsymbol{\mu}_k, \boldsymbol{\lambda}_k^{-1}\right)_{1\leq k\leq N_c} \mid (y_i, z_i)_{1\leq i\leq N}\right),$$

where the response variable is positive and real valued , i.e., $y_i \in (0, \infty)$ and the likelihood function $p\left(\left(\boldsymbol{w}_k, \boldsymbol{\mu}_k, \boldsymbol{\lambda}_k^{-1}\right)_{1\leq k\leq N_c} \mid (y_i, z_i)_{1\leq i\leq N}\right)$, is as specified in Section 1.2.1. As in the above mentioned references we include an additional hyper-parameter $\beta \in (0, \infty)$, so that the complete parameter vector is of the form

$$\boldsymbol{\theta} = \left(\left(\boldsymbol{w}_k, \boldsymbol{\mu}_k, \boldsymbol{\lambda}_k\right)_{1\leq k\leq 3}, \beta\right) \in \boldsymbol{\Delta}^3 \times \mathbb{R}^7,$$

where $\boldsymbol{\Delta}^3$ denotes the standard simplex in $\mathbb{R}^3$. We choose the prior $p_{\mathrm{prior}}(\theta)$ such that

$$\begin{aligned} \boldsymbol{\mu}_k &\sim \mathcal{N}(m, \kappa^{-1}), \\ \boldsymbol{\lambda}_k &\sim \mathrm{Gamma}(\alpha, \beta), \\ \beta &\sim \mathrm{Gamma}(g, h), \\ (\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3) &\sim \mathrm{Dirichlet}_3(1, 1, 1). \end{aligned}$$

with $m = M, \kappa = 4/R^2$, $\alpha = 2$, $g = 0.2$, $h = 100g/(\alpha R^2)$, where M and R denotes the mean and the range of the data $(y_i)_{1\leq i\leq N}$, respectively. Under this assumptions on the form of the prior, the explicit form of the resulting posterior density reads ([24, Section 2.1]):

$$\begin{aligned} p(\boldsymbol{\theta} \mid (y_i, z_i)_{1\leq i\leq N}) = \frac{\kappa^{K/2} g^h \beta^{N_c\alpha+g-1}}{\Gamma(\alpha)^K \Gamma(g)(2\pi)^{\frac{n+K}{2}}} &\left(\prod_{k=1}^{N_c} \boldsymbol{\lambda}_k\right)^{\alpha-1} \\ &\times \exp\left(-\frac{\kappa}{2}\sum_{k=1}^{N_c}(\boldsymbol{\mu}_k - M)^2 - \beta\left(h + \sum_{k=1}^{K}\boldsymbol{\lambda}_k\right)\right) \\ &\times \prod_{i=1}^{N}\left[\sum_{k=1}^{N_c} \boldsymbol{w}_k \boldsymbol{\lambda}_k^{1/2} \exp\left(-\frac{\boldsymbol{\lambda}_k}{2}(y_i - \boldsymbol{\mu}_k)^2\right)\right], \end{aligned}$$

where $\Gamma(\cdot)$ denotes the gamma function.

Remark 4.8.1. *It can be easily verified that the potential function U_{DW} satisfies the condition Assumption 7 and Assumption 8, thus we can expect (3.1-3.3) with $U = U_{DW}$ to be geometrically ergodic for a wide range of choices of the memory kernel function $\boldsymbol{K}$. However, since the domain is unbounded for this system, we can't conclude geometric ergodicity for the numerical discretisation.*[4] *For the potential function U_{GM} associated with the posterior distribution of the Bayesian Gaussian mixture model described in (ii), one can easily verify that it grows at most linearly in β and $\lambda_k, k = 1, \ldots, N_c$ when*

[4] In fact, since the potential function is a perturbed Gaussian, we expect that geometric ergodicity can still be shown for this system despite the fact that the domain is unbounded.

the respective variables tend to ∞. This observation is rather worrisome as it implies that we can not conclude geometric ergodicity by the criteria presented in Section 3.4. In particular we can not conclude the validity of a central limit theorem for this model. Results presented for this model should therefore be treated with care.

We evaluate the schemes both in terms of their performance in the computation of stationary averages and dynamical averages. More specifically, for suitable real-valued observables[5] $\varphi, \phi : \Omega_{\boldsymbol{q}} \times \Omega_{\boldsymbol{p}} \to \mathbb{R}$ we compute estimates of the expectation $\mathbb{E}_\mu \varphi$ and estimates of time auto-covariance functions associated with these observables, i.e., estimates of $\mathrm{Cov}(\varphi, \phi, \cdot) : [0, \infty) \to \mathbb{R}$, which is defined such that

$$\mathrm{Cov}(\varphi, \phi, s) = \mathbb{E}_{\boldsymbol{W}} \left[\big(\varphi(\boldsymbol{q}(0), \boldsymbol{p}(0)) - \mathbb{E}_\mu \varphi\big) \big(\varphi(\boldsymbol{q}(s), \boldsymbol{p}(s)) - \mathbb{E}_\mu \phi\big) \mid (\boldsymbol{q}(0), \boldsymbol{p}(0)) \sim \mu \right]. \tag{4.80}$$

From a numerically computed trajectory $(\boldsymbol{x}_k) = (\boldsymbol{q}_k, \boldsymbol{p}_k, \boldsymbol{s}_k)_{1 \le k \le N}$, of finite length $N \in \mathbb{N}$, we compute estimates of $\mathbb{E}_\mu \varphi$ as

$$\mu^N_{\Delta t}[\varphi] = \frac{1}{N} \sum_{k=0}^{N-1} \varphi(\boldsymbol{q}_k, \boldsymbol{p}_k),$$

where $\Delta t > 0$ denotes the step size used in the numerical scheme. Similarly, we compute estimates of the covariance function $\mathrm{Cov}(\varphi, \phi, \cdot)$ for discrete lag-times $\Delta t s, s \in \mathbb{N}, s \ll N$, as

$$\mathrm{Cov}^N_{\Delta t}(\varphi, \phi, s) = \frac{1}{N - s} \sum_{k=0}^{N-1-s} \big(\varphi(\boldsymbol{q}_k, \boldsymbol{p}_k) - \mu^N_{\Delta t}[\varphi]\big) \big(\phi(\boldsymbol{q}_{k+s}, \boldsymbol{p}_{k+s}) - \mu^N_{\Delta t}[\phi]\big). \tag{4.81}$$

An important performance criteria for the computation of stationary averages is the magnitude of the stepsize dependent systematic discretisation bias of ergodic averages

$$\mathcal{E}_{\Delta t}[\varphi] = \lim_{N \to \infty} \big| \mu^N_{\Delta t}[\varphi] - \mathbb{E}_\mu \varphi \big|, \tag{4.82}$$

which we estimate by the Monte-Carlo estimate

$$\mathcal{E}^N_{\Delta t}[\varphi] = \big| \mu^N_{\Delta t}[\varphi] - \mathbb{E}_\mu \varphi \big|.$$

In the numerical experiments presented below we either provide confidence intervals for this Monte-Carlo error or choose N sufficiently large so that the magnitude of the sampling error is negligible in comparison to the estimated quantity. Since the stationary measure μ is Gaussian in $\boldsymbol{p}$ and $\boldsymbol{s}$, we are primarily interested in the discretisation bias for observables which are functions of $\boldsymbol{q}$. A delicate question is the choice of the observable φ on whose basis we asses the discretisation error. It is not uncommon that incidental cancellation effects occur for certain combinations of a numerical scheme, a potential function, and an observable. For this reason we therefore base our analysis on an error measure which is can be considered a .That is for the one-dimensional Gibbs-distribution associated with the potential function U_{DW}, we partition the effective support $[a, b] \subset \mathbb{R}$ of the marginal measure of μ into n_B equisized bins,

$$B_i = \left[\frac{a + i(b-a)}{n_B(b-a)}, \frac{a + (i+1)(b-a)}{n_B(b-a)} \right], \ 1 \le i \le n_B,$$

[5] Note that we do note consider statistics of the of the auxiliary variable(s) $\boldsymbol{s}$ here.

and define the observables

$$\varphi_{B_i}(\boldsymbol{q}) = \mathbb{1}_{B_i}(\boldsymbol{q}) - \int_{\Omega_{\boldsymbol{x}}} \mathbb{1}_{B_i}(\boldsymbol{q})\mu(\mathrm{d}\boldsymbol{x}),\ 1 \leq i \leq n_B,$$

which measure the error in the corresponding indicator functions. As an error measure we then consider the quantity

$$\varphi_{\mathrm{MAE}}((\boldsymbol{q}_k)_{1\leq k\leq N}) := \sum_{i=1}^{n_B} \left|\mu_{\Delta t}^N\left[\varphi_{B_i}\right]\right| = \sum_{i=1}^{n_B} \left| \left(\frac{1}{N}\sum_{k=0}^{N-1} \mathbb{1}_{B_i}(\boldsymbol{q}_k) - \int_{\Omega_{\boldsymbol{x}}} \mathbb{1}_{B_i}(\boldsymbol{q})\mu(\mathrm{d}\boldsymbol{x}) \right) \right|,$$

which, following [70], we refer to as the "mean absolute error" (MAE). For multi-dimensional problems it is usually not feasible to compute the MAE. For this reason we asses discretisation bias based on the quantity

$$\mathcal{E}_{\Delta t,\mathrm{CT}}^N\left((\boldsymbol{q}_k)_{1\leq k\leq N}\right) = |\mu_{\Delta t}^N[\varphi_{CT,i}] - 1|, \tag{4.83}$$

where

$$\varphi_{CT,i}(\boldsymbol{q}) = |\boldsymbol{q}_i \partial_{\boldsymbol{q}_i} U(\boldsymbol{q}) - 1|.$$

We refer to $\mathcal{E}_{\Delta t,CT}^N$ as the error in configurational temperature (see Section 1.1.5).

Our motivation to compute dynamical averages of the form (4.80) is twofold. First, in situations where the purpose of the simulation of a discretised GLE dynamic is the computation of stationary averages, we compute estimates of the form (4.81) in order to estimate the integrated autocorrelation time of observable under the dynamics of the (stationary) Markov-chain associated with the discrete dynamics. The integrated autocorrelation time of an observable φ is defined as

$$\tau_\varphi := \lim_{s_{\max}\to\infty} \lim_{N\to\infty} \sum_{s=0}^{s_{\max}} \mathrm{Cov}_{\Delta t}^N(\varphi,\varphi,s).$$

The integrated autocorrelation time is commonly used as a benchmark for the (asymptotic) sampling efficiency of Markov chain type algorithms. Provided that a central limit theorem of the form,

$$N^{-1/2}(\mu_{\Delta t}^N[\varphi] - \mathbb{E}_\mu\varphi) \sim \mathcal{N}(0,\xi_\varphi^2),\ \text{as } N\to\infty,$$

holds, it can be shown (see e.g. [113]) that the asymptotic variance ξ_φ^2 is related to the integrated auto-correlation time τ_φ via the relation

$$\xi_\varphi^2 = \tau_\varphi \sigma_\varphi^2,$$

where $\sigma_\varphi^2 = \mathbb{E}_\mu\left[(\varphi - \mathbb{E}_\mu\varphi)^2\right]$. An estimate of the integrated auto-correlation time from a trajectory of finite length is typically computed as

$$\tau_\varphi^{N,s_{\max}} = \sum_{s=0}^{s_{\max}} \mathrm{Cov}_{\Delta t}^N(\varphi,\varphi,s).$$

The estimation of the integrated autocorrelation time τ_φ is delicate as there is a tradeoff between variance of the estimate and the bias of the estimate which both depend on the choice of $s_{\max}$. In the case of the results reported in this chapter we ensured that our

estimates are sufficiently accurate by monitoring changes in the estimate when either N and $s_{\max}$ are varied and we compared estimates of τ_φ obtained from independent simulation runs.

Another reason why we compute estimates of dynamical averages of the form (4.80) when evaluating the performance of GLE schemes is due to the fact that the very purpose to employ a GLE approach in many modelling application is exactly to obtain good approximations of dynamical quantities. We provide some preliminary results on the effect of the choice of the step size and the choice of the numerical integration scheme on the systematic bias of estimates of dynamical quantities in Section 4.8.2.

4.8.1 Comparison with previously proposed GLE schemes

We compare the MAE incurred with methods proposed in this thesis with the MAE incurred with methods previously proposed in the literature [7, 115] in the case of the simple models system described by the double-well potential U_{DW} which we specified in 4.79. The schemes proposed in [7] allow only the simulation of GLEs where the memory kernel is of the form of a finite Prony series. For this reason we choose the memory kernels in the GLE to be a positive linear combinations of two exponential functions. More specifically, we let

$$\boldsymbol{K}(t) = K_i(t) := \epsilon^{-2i} K(t\epsilon^{-i}), \tag{4.84}$$

where

$$K(t) := \sum_{k=1}^{2} c_k \exp(-\lambda_k t), \tag{4.85}$$

with

$$\begin{aligned} c_1 = \frac{5}{2},\, c_2 = \frac{1}{2}, \\ \lambda_1 = \frac{1}{4},\, \lambda_2 = \frac{1}{8}. \end{aligned} \tag{4.86}$$

Figure 4.1 shows the MAE for the methods proposed in Section 4.1. As expected all these methods show second order convergence. Moreover, we find that the ordering in terms of accuracy among the methods of the H-OU-family is very similar to what has been observed in the past in case of the corresponding splitting schemes for the white noise Langevin equation [70, 72] with the gle-BAOAB method clearly outperforming other H-OU splitting schemes. In particular, for fixed stepsize Δt, we find that using gle-BAOAB, instead of gle-OBABO results in a reduction of the MAE by a factor of up to 6 when used in combination with the memory kernel K_0. In the case of K_1 and K_2 this discrepancy is even more pronounced. In the latter setup the MAE is reduced by a factor of 18 when gle-BAOAB is used instead of gle-OBABO. In the same setup we can use an up to 10 times larger step size in gle-BAOAB and incur an error of the same magnitude as in gle-OBABO. Under the premise that the integrated autocorrelation time scales inversely with the stepsize, this corresponds to an increase in (asymptotic) efficiency by the same factor.

For moderate variance and strong autocorrelation of the noise we observe that the error incurred in the methods (B.6) and (B.7) is very similar to the error incurred in the gle-BAOAB method (Figure 4.2, a,b). The high accuracy of (B.6) is not surprising as the authors in [7] specifically design this method for the sampling of accurate configurational averages. However, in the case of (B.7), the authors do not choose

their choice of ordering in the splitting scheme as the result of a systematic analysis of the discretisation error in configurational averages. Moreover, from further numerical experiments, whose results are not displayed here, we found that the integration order used in (B.7) is optimal within the families of integrators based on the decomposition of the vector field in the SDE as described in [115] and variation of the integration sequence in all cases lead to a deterioration of the sampling accuracy in comparison to (B.7).

In the case of $K = K_2$ the maximum stable stepsize for the methods B.5, B.6 and B.7 is drastically reduced (Figure 4.2, c), while the maximum admissible step size for schemes based on an H-OU decomposition is not affected.

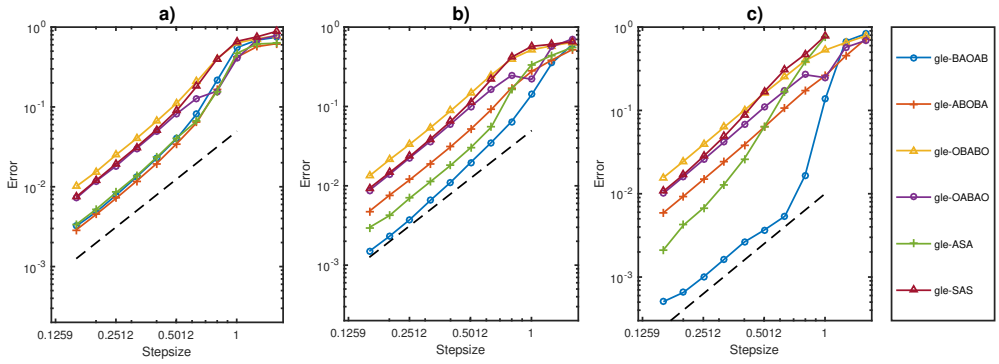

Figure 4.1: Log-Log plot of the MAE, obtained from GLE integrators proposed in this thesis which were applied to a GLE with potential function (4.79). The results for $\boldsymbol{K} = K_1, \boldsymbol{K} = K_2$, and $\boldsymbol{K} = K_3$, where K_i with $i = 1, 2, 3$ as defined in (4.84), are shown in panel a), b) and c), respectively. In total, 100 trajectories, all initialised in accordance with the exact equilibrium distribution of the extended system, were simulated over a physical time period of length 10.000.000 to obtain the statistics. Any missing error value indicates numerical instability of the respective method for the corresponding stepsize. The dashed black line corresponds to a second order decay.

4.8.2 Discretisation bias in dynamical observables

We provide some preliminary results regarding the approximation properties of gle-BAOAB for dynamical observables. Using the same setup as for the simulation described in Section 4.8.1 with $\boldsymbol{K} = K_2$, we compute the auto-correlation function of the position $\mathrm{Cov}(\boldsymbol{q}, \boldsymbol{q}, \cdot)$ and the momentum $\mathrm{Cov}(\boldsymbol{p}, \boldsymbol{p}, \cdot)$ from a finite trajectory of length $N = \lceil 1000.000/\Delta t \rceil$, as specified in (4.81). We choose the stepsize relatively large as $\Delta t = 0.6355$. We compare the obtained autocorrelation functions with

(i) the auto-correlation function obtained from a simulation gle-OBABO, which was computed using the same stepsize $\Delta t = 0.6355$.

(ii) a reference solution of auto-correlation function, which was computed using a significantly smaller step size $\Delta t = 0.0318$. We verified that for this stepsize the auto-correlation function computed by either gle-BAOAB were gle-OBABO were (at least visually) indistinguishable.

The results are shown in Figure 4.3. We find that both the momentum auto-correation function and the position auto-correlation function are significantly better approximated by gle-BAOAB.

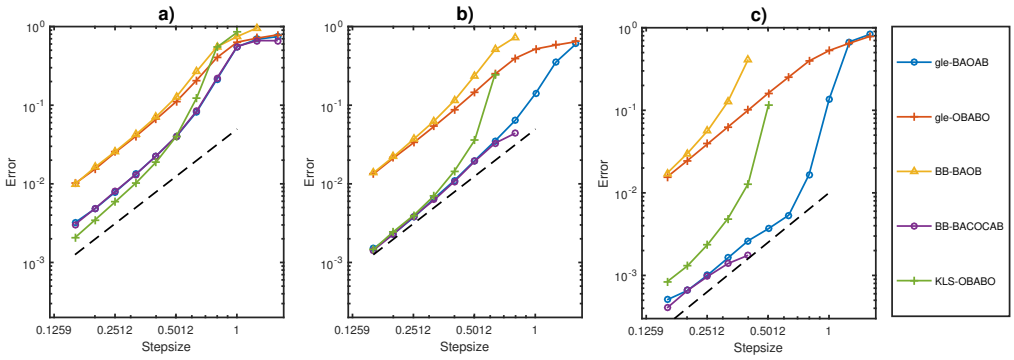

Figure 4.2: Log-Log plot of stepsize vs MAE. obtained from either gle-BAOAB or integrators which were previously proposed in the literature. Again, we consider a GLE with potential function (4.79). The results for $\boldsymbol{K} = K_1, \boldsymbol{K} = K_2$, and $\boldsymbol{K} = K_3$, where K_i with $i = 1, 2, 3$ as defined in (4.84), are shown in panel a),b) and c), respectively. In total, 100 trajectories, all initialised in accordance with the exact equilibrium distribution of the extended system, were simulated over a physical time period of length 10.000.000 to obtain the statistics. Any missing error value indicates numerical instability of the respective method for the corresponding stepsize. The dashed black line corresponds to a second order decay.

4.8.3 Parameter dependent accuracy of gle-BAOAB

In order to support the results derived by the singular perturbation ansatz in Section 4.6.2, we evaluate the sampling accuracy of gle-BAOAB when applied to a GLE with a simple exponentially decaying memory kernel, i.e.,

$$\boldsymbol{K}(t) = \gamma e^{-t/\tau}, \gamma > 0, \tau > 0, \tag{4.87}$$

and the potential function (4.79). As predicted we find in the overdamped limit, i.e., for a scaling of the form

$$\boldsymbol{K}(t) = K_i(t) := \epsilon^{-2i} K(t\epsilon^{-i}), \tag{4.88}$$

a decrease of the MAE with increasing index i (See Figure 4.4 b). Moreover, for parameter values $\lambda = 128, \tau = 1/16$, we find the predicted 4th order decay of the discretisation bias as Δt tends to 0. For the chosen range of parameter values we further observe

(i) a decrease of the MAE in the white noise limit $K_i(t) := \epsilon^{-i} K(t\epsilon^{-i})$, as i increases (Figure 4.4 **a**).)

(ii) a decrease of the MAE for fixed decay rate $\tau = 1$ as the pre-factor γ increases (Figure 4.4 **d**).)

(iii) no systematic change of the magnitude of the MAE for fixed pre-factor $\gamma = 4$ and varying decay rate τ (Figure 4.4 **c**).)

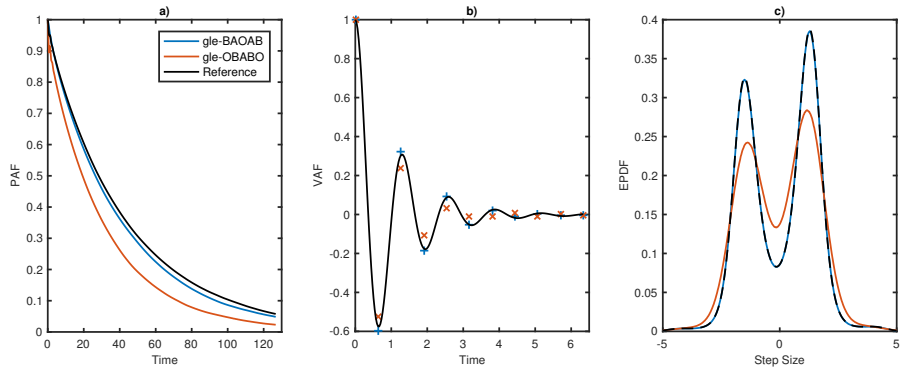

Figure 4.3: Plots showing the position autocorrelation function (a), velocity autocorrelation function (b), and invariant density (c) for the GLE with potential function (4.79) and memory kernel K_2 as defined in (4.84) obtained by either using gle-BAOAB (blue) or gle-OBABO (red) with a fixed stepsize, $\Delta t = 0.6355$. Statistics were computed from a single trajectory of length 1000.000 (physical time). The reference solution in plot a) and b) were computed using a reduced stepsize of $\Delta t = 0.0318$, which was verified to be sufficiently small for the respective autocorrelation functions when computed with either integrator to be visually indistinguishable. The dashed black line corresponds to a second order decay.

4.8.4 Gaussian mixture model

In [22] it was proposed to use quasi-Markovian GLEs in order to improve sampling of ill conditioned (typically uni-modal) target densities[6] In this section we combine the approach proposed in the above mentioned reference with the gle-BAOAB integration scheme and compare the performance of the resulting sampling scheme with

(i) a BAOAB discretisation of the underdamped Langevin equation.

(ii) a gle-OBABO discretisation of the same GLE dynamics, which we use in gle-BAOAB. This combination of GLE dynamics and discretisation scheme corresponds to the sampling scheme proposed in [22].

As a benchmark system we consider the Bayesian Gaussian mixture model specified in item (ii). We parametrise both the gle-BAOAB scheme and the gle-OBABO scheme with the pre-optimised memory kernel kv-8-8 obtained from the website GLE4MD [20]. This kernel is optimised for the frequency range $[\omega_{min}, \omega_{max}] = [0.0001, 10000]$. We provide the exact parametrisation in appendix C. We compare the performance of the sampling schemes

(i) in terms of the observed discretisation bias which we measure by the error in configurational temperature $\mathcal{E}^N_{\Delta t,\mathrm{CT}}$ as defined in (4.83),

(ii) in terms of an estimate of the autocorrelation times τ_{P_i}, where $P_i : \boldsymbol{q} \mapsto \boldsymbol{q}_i$, denotes the projection operator which selects the i-th component of $\boldsymbol{q}$.

For ld-BAOAB we report two simulation runs. The simulation run corresponding to the parameter values $\gamma = 1.0, \Delta t = .01$ was obtained as the result of minimising the

[6]We refer to a uni-modal density as ill conditioned iff its' associated covariance matrix is ill conditioned.

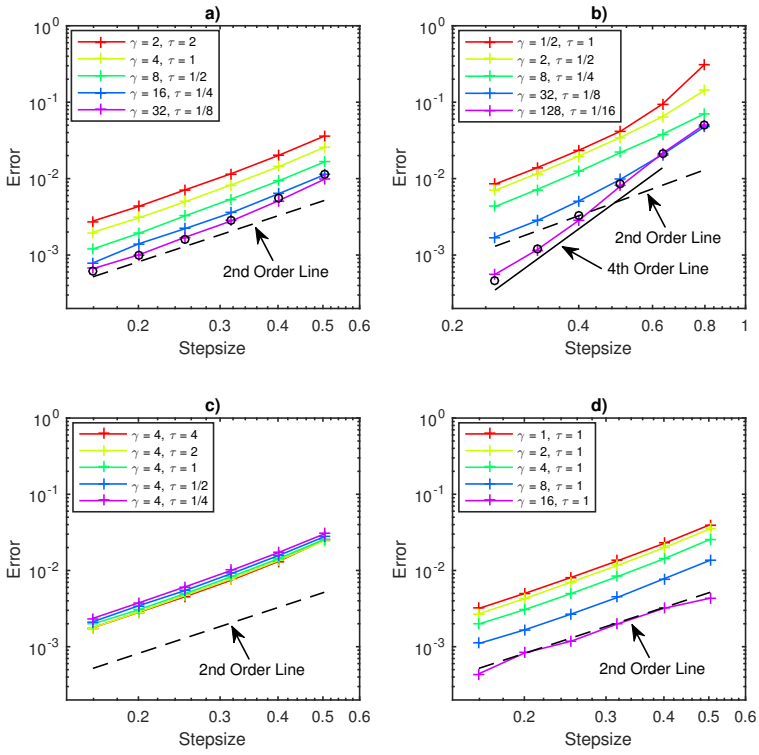

Figure 4.4: Log-Log plot of stepsize vs mean absolute error in distribution of the invariant marginal density in q using the numerical integrator gle-BAOAB applied to the GLE with potential function (4.79) and memory kernel $K(t) = \gamma e^{-t/\tau}$ with varying values of γ and τ. Black circles show the observed MAE of the respective limiting dynamics.

integrated autocorrelation time for the slowest parameters by varying the stepsize after fixing the friction coefficient to $\gamma = 1.0$. The simulation run corresponding to the parameter values $\gamma = .1$ with $\Delta t = .01$ was obtained as the result of minimising the integrated autocorrelation time for the "slowest paramete"[7] by simultaneously optimising both the stepsize as well as the friction coefficient γ. The results reported for gle-BAOAB and gle-OBABO use a stepsize $\Delta t = .02$, which was determined approximately as the maximum admissible stepsize with a few (short!) test runs. We find that in terms of sampling efficiency which we measured in terms of the integrated autocorrelation time of the "slowest" sampled parameter λ_1, the GLE schemes clearly outperform these as Figure 4.5 shows. Between the GLE schemes we find that the discretisation error in the sample obtained from gle-BAOAB is significantly smaller than the discretisation error in the sample obtained with gle-OBABO. The improvement in terms of the maximum admissible stable stepsize of the GLE methods in comparison to the Langevin schemes is an interesting feature. Presumably, this is due to resonance effects which occur in the discretised dynamics of the underdamped Langevin due to an insufficient damping of fast frequency modes for the tuned value of the friction coefficient.

[7]With that we mean the parameter with the largest integrated autocorrelation time as shown in Figure 4.5

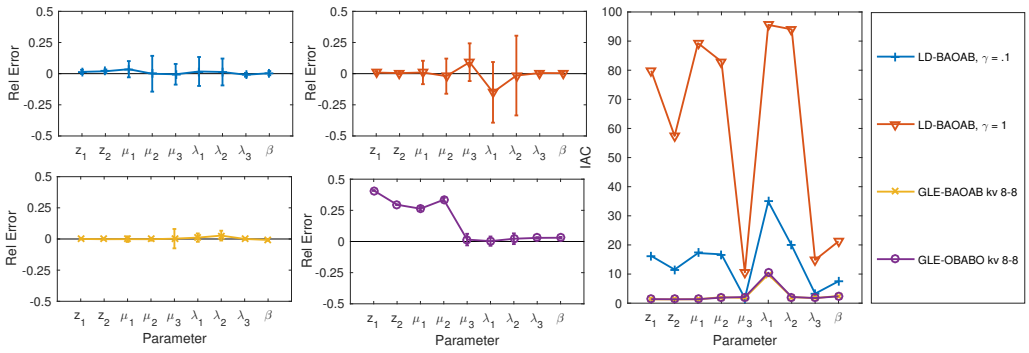

Figure 4.5: Sampling statistics for the numerical experiments performed on the Hidalgo stamp dataset. The rightmost plot shows the integrated auto-corrleation time (labeled as IAC) for each sampled parameter of the Gaussian mixture model. The remaining plots on the left shows the discretisation bias for each method separately. The discretisation bias is measured for each sampled parameter separately as the deviation of the empirical mean of the observable (1.13) from the exact mean value under the target measure.

Chapter 5

Adaptive generalized Langevin dynamics

In many practical contexts, the gradient can be expensive or intractable to compute; as a result, there has been a recent surge in the development of MCMC algorithms that require only unbiased estimators of gradient information [125, 97, 23, 27, 111]. Algorithms of this type, such as Stochastic Gradient Langevin Dynamics (SGLD) perform impressively on Bayesian inference tasks involving highly expressive models with large quantities of training data [8]. The theoretical understanding of SGLD has also been advanced in several recent works [120, 124].

A standard feature of SGLD schemes is the assumption that the errors in the gradient estimates used in the Markov chain are statistically independent. However, correlated samples arise naturally in many situations, for example when the minibatches of data used to estimate the gradients at successive timesteps are not independently drawn from the population dataset. A particular case is where the data points are distributed across several machines [3, 4], or when minibatches are only partially refreshed between Markov chain time steps.

In this chapter, we introduce an adaptive generalized dynamics (AdGLD) algorithm, which is able to deal robustly with correlated gradient noise. Our starting point in the design of AdGLD is a quasi-Markovain generalised Langevin equation of the form (1.30), i.e.,

$$\begin{aligned} \dot{\boldsymbol{\theta}}(t) &= \boldsymbol{p}(t), \\ \dot{\boldsymbol{p}}(t) &= \mathscr{G}(\boldsymbol{\theta}(t)) - \int_0^t \boldsymbol{K}(t-s)\boldsymbol{p}(s)\mathrm{d}s + \boldsymbol{\eta}(t). \end{aligned} \tag{5.1}$$

where $\mathscr{G}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \log p(\boldsymbol{\theta}|x_{1:N})$ is the gradient of a posterior density and the memory kernel $\boldsymbol{K}$ and the noise process $(\boldsymbol{\eta}(t))_{t>0}$ are as defined in Section 1.3.

The method described in this chapter is an "adaptive" variant of the GLE in which the strength of the dissipation term is automatically adjusted to match the a priori unknown amplitude of the stochastic process $(\boldsymbol{\eta}(t))_{t>0}$ so as to maintain invariance of the target distribution. Basing the sampling procedure on the GLE allows for the explicit modelling of a variety of partial minibatch refreshment techniques (see Section 5.1.2), including methods of online inference, distributed MCMC, and partial refreshment of minibatches in a batch inference setting. This chapter adds to the emerging literature on Bayesian sampling methods employing thermostat techniques [111, 77], which are able to achieve enhanced numerical stability properties by borrowing ideas from the mature field of molecular dynamics.

The rest of this chapter is structured as follows. In Section 5.1 we review the existing literature on stochastic gradient MCMC methods, and introduce the notion of correlated gradient estimates in MCMC algorithms. In Section 5.2, we provide the theoretical details of a novel thermostat method which allows for the construction of efficient MCMC samplers using correlated gradient estimators. In Section 5.3, we compare the performance of AdGLD to that of traditional MCMC schemes for synthetic and naturally arising inference problems.

5.1 Stochastic gradient MCMC and correlated gradient estimates

5.1.1 Stochastic Gradient Methods

In the Bayesian inference context the computational bottlenecks in the Metropolis-Hastings algorithm (Algorithm 1) for proposals of the form (1.22) typically are the evaluation of the acceptance ratio $p_{acc}(\widetilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$ and the gradient of the log-likelihood, $\mathscr{G}(\boldsymbol{\theta})$. Stochastic gradient Monte Carlo methods seek to eliminate these bottlenecks in two ways: (i) the acceptance ratio $\alpha(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta})$ is set to 1, so that the acceptance step may effectively be ignored, and the evaluation of the target density is bypassed; and (ii) the gradient term $\nabla \log p(\boldsymbol{\theta}|x_{1:N})$ in the proposal is replaced with an unbiased estimator, so that successive samples $\boldsymbol{\theta}_t, t = 0, 1, 2 \dots$ are generated by

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \frac{\varepsilon}{2}\widehat{\mathscr{G}}_t + \boldsymbol{\eta}_t^{\varepsilon} \tag{5.2}$$

where the gradient estimator typically has the form of a subsampling estimator:

$$\widehat{\mathscr{G}}_t := \widehat{\nabla}\boldsymbol{\theta}^{(t)} \log p(\boldsymbol{\theta}|x_{1:N}) = \nabla_{\boldsymbol{\theta}} \log \pi(\boldsymbol{\theta}) + \frac{N}{|I_t|}\sum_{i \in I_t} \nabla_{\boldsymbol{\theta}} \log p(x_i|\boldsymbol{\theta}) \tag{5.3}$$

where the random subset $I_t \subset \{1, \dots, N\}$ is the minibatch of data used to construct the gradient estimate, reducing the computational burden of simulating from the proposal from $O(N)$ to $o(N)$. The index sets $(I_t)_{t\in\mathbb{N}}$ are typically taken to be independent for each time step.

5.1.2 Correlated Stochastic Gradients

A principal contribution of this work is to extend the stochastic gradient MCMC methods discussed in Section 5.1.1 for use in situations where the gradient estimators (5.3) are temporally correlated (i.e. when the minibatch index sets $(I_t)_{t\in\mathbb{N}}$ are not independent). Such scenarios are of practical interest when it is infeasible to draw independent minibatches at each Markov chain time step, or when computational savings might be achievable by reducing the rate of minibatch refreshment. Theoretically, we connect our methods with SDEs with temporally correlated stochastic forcing terms, and introduce a new thermostat method for stable numerical simulation.

In the remainder of this section, we introduce distributions over the index sets $(I_t)_{t\in\mathbb{N}}$ that model correlation in the gradient estimators (5.3).

k-FIFO

The $k-$FIFO replacement scheme corresponds to first-in, first-out queue of data points. Between each time step of the Markov chain, k data points leave the queue, and k new

data points, drawn uniformly and independently from the population $x_{1:N}$ join the queue. The data points present in the queue are used to form the subsampled gradient estimator at that Markov chain time point. More precisely, let $(s_n)_{n\in\mathbb{N}}$ be a stream of independent random data points drawn uniformly from the population $x_{1:N}$. For minibatches of size n, we then set $I_t = \{s_{(t-1)k+1}, \ldots, s_{((t-1)k+n+1)}\}$. The $k-$FIFO scheme yields the following covariance kernel for gradient estimators:

$$\mathrm{Cov}(\widehat{\mathscr{G}_t}, \widehat{\mathscr{G}_{t+m}}) = \boldsymbol{\sigma}^2 \max\left(0, 1 - \frac{km}{n}\right), \tag{5.4}$$

where $\boldsymbol{\sigma}^2 = \mathrm{Var}(\widehat{\mathscr{G}_t})$. In general, $\mathrm{Var}(\widehat{\mathscr{G}_t})$ may dependent on the value of the parameter $\boldsymbol{\theta}$, so that $\boldsymbol{\sigma}^2$ is a function of $\boldsymbol{\theta}$, however throughout this chapter we will assume that the variance of the gradient noise is constant in $\boldsymbol{\theta}$. Although theoretically not well justified this assumption is commonly made in the context of stochastic gradient sampling algorithms; (see e.g. [27, 74]). In the numerical experiments we performed we did indeed not observe any noticeable adverse effect on sampling accuracy caused by this assumption; see Section 5.3.

k-Random Replacement

Rather than eliminating the "oldest" data points in the minibatch at each partial refreshment, we may choose k data points to remove uniformly at random. This leads to the k-random replacement (k-RR) scheme. The k-RR scheme yields the following covariance kernel for gradient estimators

$$\mathrm{Cov}(\widehat{\mathscr{G}_t}, \widehat{\mathscr{G}_{t+m}}) = \boldsymbol{\sigma}^2 \left(1 - \frac{k}{n}\right)^m, \tag{5.5}$$

where $\boldsymbol{\sigma}^2 = \mathrm{Var}(\widehat{\mathscr{G}_t})$.

Poisson Random Replacement

Further developing the random replacement scheme, we can opt to refresh our data points according to some stochastic schedule. Here, we consider refreshing data point independently at each time step with some probability p; this yields a memory kernel of the form

$$\begin{aligned}
&\mathrm{Cov}(\widehat{\mathscr{G}_t}, \widehat{\mathscr{G}_{t+m}}) \\
&= \boldsymbol{\sigma}^2 \left(\sum_{k=0}^{n} \mathbb{P}(k \text{ refreshments }) \left(1 - \frac{k}{n}\right)\right)^m \\
&= \boldsymbol{\sigma}^2 (1-p)^m,
\end{aligned} \tag{5.6}$$

where $\boldsymbol{\sigma}^2 = \mathrm{Var}(\widehat{\mathscr{G}_t})$. We refer to this replacement rule as Poisson random replacement (k-PRR). This scheme can be viewed as a discretisation of a continuous scheme in which data points are refreshed according to the increments of a Poisson process; it has the attractive property that p may be tuned to set the expected number of steps in which no refreshment at all is required.

5.1.3 Application of thermostat methods in noisy gradient systems

The additional noise present in stochastic gradient Monte Carlo methods due to the gradient estimator (5.3) can significantly affect the dynamics of the chain; this has been studied in detail from a theoretical perspective [124, 120], and new MCMC routines have been introduced to stabilize the dynamics in the presence of this additional (unknown) noise [111, 77, 27, 74], using thermostat-based constructions from the field of molecular dynamics; see [74].

Introducing correlation into these gradient estimators can further disrupt the dynamics of these stochastic processes; see Section 5.3 for empirical confirmation. In the following section we introduce a new thermostat method based on the generalized Langevin equation designed specifically to deal with the addition of correlated noise into these stochastic processes which we term the covariance controlled adaptive generalized Langevin dynamics sampler (CCAdGLD).

5.2 Adaptive thermostats for time correlated gradient perturbations

We introduce the methods proposed in this chapter in an SDE formulation. We therefore treat in the following the time index t as a continuous variable, i.e., $t \in \mathbb{R}$. We also decompose the gradient estimate into a clean gradient and a noise process $(\boldsymbol{\eta}(t))_{t>0}$:

$$\widehat{\mathscr{G}_t} = \mathscr{G}_t + \boldsymbol{\eta}(t),$$

which we assume to be Gaussian with an unknown, potentially parameter dependent auto-covariance function. We next describe how the GLE can be combined with thermostat methods to allow accurate sampling of the posterior $p(\boldsymbol{\theta}|x_{1:N})$ in the situation where only a perturbed estimate $\mathscr{G}_t$ is available.

5.2.1 The general framework

When the form of the auto-covariance function of the noise process $(\boldsymbol{\eta}(t))_{t>0}$ leads to an ergodic dynamic of the GLE (1.28), one can expect that if the noisy gradient $\mathscr{G}_t$ is embedded within a GLE with an appropriate convolution term that the resulting stochastic dynamic would lead to an ergodic process which would allow sampling of $p(\boldsymbol{\theta}|x_{1:N})$. However, since the exact form of the auto-covariance function of the noise is unknown one is instead bound to derive an appropriate dynamic from a generic stochastic integro-differential equation:

$$\begin{aligned} \dot{\boldsymbol{\theta}}(t) &= \boldsymbol{p}(t), \\ \dot{\boldsymbol{p}}(t) &= \mathscr{G}(\boldsymbol{\theta}(t)) + \boldsymbol{\eta}(t) - \int_0^t \widehat{\boldsymbol{K}}^{(t)}(t-s)\boldsymbol{p}(s)\mathrm{d}s, \end{aligned} \tag{5.7}$$

which resembles the GLE with the exception that the auto-covariance function of the noise in the convolution term is substituted by a time dependent estimator $\widehat{\boldsymbol{K}}^{(t)} \approx \boldsymbol{K}$.

Markovian form of the noise process

In order to make the estimation of the memory kernel $\widehat{\boldsymbol{K}}^{(t)}$ tractable, we assume that $\boldsymbol{\eta}$ is a stationary Gaussian process which is representable in the Markovian form intro-

duced in Section 3.2, i.e., there are suitable matrices

$$\boldsymbol{\Gamma}_{1,1}, \boldsymbol{\Sigma}_{1,1}, \boldsymbol{\Gamma}_{1,2}, \boldsymbol{\Gamma}_{2,1}^T, \boldsymbol{\Gamma}_{2,2}, \boldsymbol{\Sigma}_{2,2} \in \mathbb{R}^{n\times n}, \tag{5.8}$$

so that

$$\boldsymbol{\eta}(t) = \boldsymbol{\eta}_w(t) + \boldsymbol{\eta}_c(t),$$

where

$$\boldsymbol{\eta}_w(t) = \boldsymbol{\Sigma}_{1,1}\dot{\boldsymbol{W}}_1(t),$$

and

$$\boldsymbol{\eta}_c = -\boldsymbol{\Gamma}_{1,2}\boldsymbol{\eta}_c',$$

where $\boldsymbol{\eta}_c'$ is the solution of

$$\dot{\boldsymbol{\eta}}_c' = -\boldsymbol{\Gamma}_{2,2}\boldsymbol{\eta}_c' + \boldsymbol{\Sigma}_{2,2}\dot{\boldsymbol{W}}_2, \tag{5.9}$$

and

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_1 \\ \boldsymbol{W}_2 \end{pmatrix},$$

denotes a standard Wiener process in $\mathbb{R}^{2n}$ with independent components. One can easily verify that if the form of the matrices (5.8) is known, then one can rewrite (5.1) as

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= \mathscr{G}(\boldsymbol{\theta}) - \boldsymbol{\Gamma}_{1,2}\boldsymbol{g} - \boldsymbol{\Gamma}_{1,1}\boldsymbol{p} + \boldsymbol{\eta}_w + \boldsymbol{\eta}_c, \\ \dot{\boldsymbol{g}} &= \boldsymbol{\Gamma}_{2,1}\boldsymbol{p} - \boldsymbol{\Gamma}_{2,2}\boldsymbol{g}, \end{aligned} \tag{5.10}$$

Consequentially, the problem of constructing a suitable estimate $\widehat{\boldsymbol{K}}^{(t)}$ of $\boldsymbol{K}$ is equivalent to the estimation of the forms of the respective matrices (5.8). Moreover, if the noise $\boldsymbol{\eta}$ is induced by one of the replacement rules introduced in Section 5.1.2, the estimation problem is further simplified since the auto-correlation function associated with each of these replacement rules is known a priori. In terms of (5.10) this means that the matrices $\boldsymbol{\Gamma}_{2,2}$ and $\boldsymbol{\Sigma}_{2,2}$ can be assumed to be known. More specifically, for the random replacement rule and the Poisson replacement rule it follows that

$$\boldsymbol{\Gamma}_{2,2} = \lambda \boldsymbol{I}_n, \boldsymbol{\Sigma}_{2,2} = \sqrt{2\lambda}\boldsymbol{I}_n, \text{ with } \lambda > 0, \tag{5.11}$$

as well as

$$\boldsymbol{\Gamma}_{1,1} = \boldsymbol{0}, \tag{5.12}$$

and $\boldsymbol{\eta}_w \equiv 0$. We note that the auto-correlation function associated with the k-FIFO rule can not exactly be represented in the extended variable formalism (5.10). Instead, one would need to approximate the auto-correlation function of k-FIFO by a function which can be represented in the above extended variable formalism; (see e.g. [7] where the approximation of power-law functions by finite Prony-series is considered). We do not further follow up on this issue here and instead focus on the cases of a random replacement rule or a Poisson replacement rule. In these cases we still require a suitable estimation procedure for the matrices $\boldsymbol{\Gamma}_{1,2}$ and $\boldsymbol{\Gamma}_{2,1}$. By combining the variables $\boldsymbol{\eta}_c'$ and

$\boldsymbol{g}$ as $\boldsymbol{s} = \boldsymbol{g} + \boldsymbol{\eta}_c'$ we obtain

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= \mathscr{G} - \boldsymbol{\Gamma}_{1,1}\boldsymbol{p} - \boldsymbol{\Gamma}_{1,2}\boldsymbol{s} + \boldsymbol{\Sigma}_{1,1}\dot{\boldsymbol{W}}_1, \\ \dot{\boldsymbol{s}} &= -\boldsymbol{\Gamma}_{2,1}\boldsymbol{p} - \boldsymbol{\Gamma}_{2,2}\boldsymbol{s} + \boldsymbol{\Sigma}_{2,2}\dot{\boldsymbol{W}}_2. \end{aligned} \tag{5.13}$$

Since for a given realisation of the Wiener process $\boldsymbol{W}$ the solution of (5.10) and (5.13) are identical in $\boldsymbol{\theta}$, it follows that in order to ensure that $\boldsymbol{\theta}$ as a solution of (5.10) is ergodic with invariant measure $p(\boldsymbol{\theta}|x_{1:N})$, it is sufficient to choose the matrices (5.8) such that (5.13) is ergodic. We derived sufficient conditions for geometric ergodicity of (5.13) in Section 3.4. In particular, for $\boldsymbol{\Gamma}_{2,2} = \lambda \boldsymbol{I}_n, \boldsymbol{\Sigma}_{2,2} = \sqrt{2\lambda}\boldsymbol{I}_n$ it follows by Lemma 3.1.1 that

$$\boldsymbol{\Gamma}_{1,2} = -\boldsymbol{\Gamma}_{2,1}^T \tag{5.14}$$

is necessary in order for (5.13) to possess an invariant measure of with density

$$\rho_{\boldsymbol{I},1}(\boldsymbol{\theta}, \boldsymbol{p}, \boldsymbol{s}) \propto p(\boldsymbol{\theta}|x_{1:N}) e^{-\|\boldsymbol{p}\|_2^2 - \|\boldsymbol{s}\|_2^2}.$$

If (5.11) and (5.14) holds, one can easily verify that

$$\boldsymbol{\Gamma}_{1,2}\boldsymbol{\Gamma}_{1,2}^T = \mathrm{Var}(\boldsymbol{\eta}_c(t)) = \mathrm{Var}(\mathscr{G}(\boldsymbol{\theta}(t))) = \boldsymbol{\sigma}^2.$$

Which means that we can compute estimates for $\boldsymbol{\Gamma}_{1,2}$ from estimates of the covariance matrix of the gradient noise. We describe in Section 5.2.2 how such estimates $\widehat{\boldsymbol{\sigma}}_{(t)}^2$ of $\boldsymbol{\sigma}^2$ can be obtained.

Computing estimates of $\boldsymbol{\Gamma}_{1,2}$ as Cholesky decompositions of $\widehat{\boldsymbol{\sigma}}_{(t)}^2$ is computationally costly and would raise further questions regarding the numerical stability of the method. If (5.11) and (5.14) hold one can easily verify that (5.10) can be rewrite as

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \boldsymbol{p}, \\ \dot{\boldsymbol{p}} &= \mathscr{G}(\boldsymbol{\theta}) - \boldsymbol{\sigma}^2\boldsymbol{r} - \boldsymbol{\Gamma}_{1,1}\boldsymbol{p} + \boldsymbol{\eta}_w + \boldsymbol{\eta}_c, \\ \dot{\boldsymbol{r}} &= \boldsymbol{p} - \boldsymbol{\Gamma}_{2,2}\boldsymbol{r}, \end{aligned} \tag{5.15}$$

with $\boldsymbol{r} = \boldsymbol{\Gamma}_{1,2}^T\boldsymbol{g}$. This means, that by using (5.15) instead of (5.10) as the basis for a numerical method we can avoid computing the Cholesky decompositon of the estimate of $\boldsymbol{\sigma}^2$.

5.2.2 Estimates using empirical covariance information

If gradient estimates come from subsampling, or more generally if the gradient is a weighted sum of unbiased gradient estimates, an unbiased estimate of $\boldsymbol{\sigma}^2$ can be obtained by computing the empirical covariance

$$V^{(t)} = \frac{1}{|I_t| - 1} \sum_{i \in I_t} D_i^{(t)}(\boldsymbol{\theta}) D_i^{(t)}(\boldsymbol{\theta})^\top,$$

where

$$D_i^{(t)}(\boldsymbol{\theta}) = \left(N \nabla_{\boldsymbol{\theta}} \log p(x_i|\boldsymbol{\theta}) - \widehat{\mathscr{G}_t} \right).$$

In implementation, this would be done at discrete timepoints $\mathbb{N}\Delta t \subset [0,\infty)$, where Δt corresponds to the stepsize of the discrete dynamics, hence in the continuous formulation, one would model $\widehat{\boldsymbol{\sigma}}^2_{(t)}$ as a càdlàg function i.e. $\widehat{\boldsymbol{\sigma}}^2_{(t)} = V^{(h(t))}(\boldsymbol{\theta}_{h(t)})$ with $h(t) = t-(t \mod \Delta t)$. More generally, in order to reduce the variance of the estimator $\widehat{\boldsymbol{\sigma}}^2_{(t)}$, estimates of the empirical covariance can be averaged as

$$\widehat{\boldsymbol{\sigma}}^2_{(t)} = (1-\alpha_{n(t)})\widehat{\boldsymbol{\sigma}}^2_{(h(t)-\Delta t)} + \alpha_{n(t)} V^{(h(t))}(\boldsymbol{\theta}_{h(t)}). \tag{5.16}$$

where $n(t) = h(t)/\Delta t$ and $\alpha_k \in (0,1]$. Since we assumed that the gradient noise is constant in $\boldsymbol{\theta}$ we can use a harmonic weighting, $\alpha_k = k^{-1}$. This ensures that $\widehat{\boldsymbol{\sigma}}^2_{(t)}$ is a strongly consistent estimator of $\boldsymbol{\sigma}^2$, i.e.,

$$\lim_{t\to\infty} \widehat{\boldsymbol{\sigma}}^2_{(t)} = \lim_{L\to\infty} \frac{1}{L}\sum_{k=1}^{L} V^{(k\Delta t)}(\boldsymbol{\theta}_{k\Delta t}) =: \boldsymbol{\sigma}^2, \text{ a.s.}$$

5.2.3 Covariance controlled adaptive generalized Langevin dynamics

We construct a thermostat method which is designed to correctly dissipate gradient noise induced by the replacement rules k-RR and k-PRR as a combination of the SDE dynamics (5.15), the estimation procedure of the covariance matrix $\widehat{\boldsymbol{\sigma}}^2_{(t)}$ as described in Section 5.2.2, and an additional Nosé-Hoover control, i.e.,

$$\begin{aligned}
\dot{\boldsymbol{\theta}} &= \boldsymbol{p},\\
\dot{\boldsymbol{p}} &= \mathscr{G}(\boldsymbol{\theta}) - \widehat{\boldsymbol{\sigma}}^2_{(t)}\boldsymbol{r} - \boldsymbol{\xi}\odot\boldsymbol{p} + \boldsymbol{\eta},\\
\dot{\boldsymbol{r}} &= \boldsymbol{p} - \lambda\boldsymbol{r},\\
\dot{\boldsymbol{\xi}} &= [\boldsymbol{p}\odot\boldsymbol{p} - \mathbf{1}]\mu^{-1},
\end{aligned} \tag{5.17}$$

where $\mu > 0$ is a positive constant, $\boldsymbol{\xi}(t) \in \mathbb{R}^n$, and the operator $\odot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is to be understood as a component-wise multiplication. The differential equation (5.17) corresponds up to the additonal Nosé-Hoover control to (5.15) with $\boldsymbol{\sigma}^2$ being substituted by the estimate $\widehat{\boldsymbol{\sigma}}^2_{(t)}$. The differential equation (5.17) can be considered as a non-Markovian generalisation of a thermostat method which was previously proposed in [111] under the name of covariance controlled adaptive Langevin dynamics (CCAdLD). Due to this structural resemblance we refer to the dynamics defined by (5.17) as covariance controlled adaptive generalised Langevin dynamics (CCAdGLD). The reason why we incorporate the additional Nosé-Hoover control is to achieve a more robust behaviour of the discretisation of (5.17); See Section 5.2.5.

5.2.4 Ergodic properties of CCAdGLD

The presence of the time dependent estimate $\widehat{\boldsymbol{\sigma}}^2_{(t)}$ and the additional Nosé-Hoover control makes the analysis of the ergodic properties of CCAdGLD. Provide that $\boldsymbol{\xi}$ is initialised close enough to its' equilibrium value 0, we might expect (5.17) to have similar ergodic properties as the corresponding GLE dynamics (5.15) for sufficiently large t; (for large t we can expect the value of $\widehat{\boldsymbol{\sigma}}^2_{(t)}$ to be close to $\boldsymbol{\sigma}^2$). However, proving such a result is not straightforward. While the presence of the additional Nosé-Hoover control makes the construction of a suitable Lyapunov difficult, the noisy estimate $\widehat{\boldsymbol{\sigma}}^2_{(t)}$ would require us to employ techniques which are commonly used in proofs of convergence of adaptive MCMC methods (see [104, 6]). We defer a rigorous analysis to future work.

5.2.5 Numerical discretization

We construct the numerical discretization of CCAdGLD as a symmetric splitting method along the lines of the popular BAOAB scheme [70] for Langevin dynamics. More specifically, we construct the numerical discretisation as a splitting scheme based on the following decomposition of (5.17),

$$\begin{bmatrix}\dot{\boldsymbol{\theta}}\\ \dot{\boldsymbol{p}}\\ \dot{\boldsymbol{r}}\\ \dot{\boldsymbol{\xi}}\end{bmatrix} = \underbrace{\begin{bmatrix}\boldsymbol{p}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}}_{=:\mathrm{A}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ \mathscr{G}(\boldsymbol{\theta})+\boldsymbol{\eta}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}}_{=:\mathrm{B}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ -\widehat{\boldsymbol{\sigma}}^2_{(t)}\boldsymbol{r}-\boldsymbol{\xi}\odot\boldsymbol{p}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}}_{=:\mathrm{C}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \boldsymbol{p}-\lambda\boldsymbol{r}\\ \mathbf{0}\end{bmatrix}}_{=:\mathrm{D}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\\ [\boldsymbol{p}\odot\boldsymbol{p}-\mathbf{1}]\mu^{-1}\end{bmatrix}}_{=:\mathrm{E}}. \tag{5.18}$$

Based on this decomposition we construct a stochastic splitting scheme using the integration sequence BAEDCDEAB. We provide an implementation of the resulting scheme in algorithm 6. From the results provided in [18, 1, 72], we conclude that the systematic error in ergodic averages introduced by the time discretization is at least $O(\Delta t)$.[1] From a detailed analysis of algorithm 6 and observations in numerical experiments we find that for large $\lambda > 0$, (i.e., for a fast decaying auto-correlation function of the noise process), dissipation through the auxiliary variable r becomes ineffective. This is the reason why we include the additional Nosé-Hoover control in the formulation of CCAdGLD. While for small $\lambda > 0$ the Nosé-Hoover control does not have any noticable adverse affect on the accuracy of the sampling it functions as an adaptive thermostat (see [58]) in the situation where the gradient noise is close to being Markovian. This effect makes the performance of the sampling method less sensitive to the value of λ (in comparison to a similar schemes which would not incorporate a Nosé-Hoover control).

Algorithm 6 Numerical Implementation of CCAdGLD as a symmetric splitting scheme

Require: $\boldsymbol{\theta}_0, \boldsymbol{p}_0, \boldsymbol{r}_0, \boldsymbol{\xi}_0, \mathcal{I}_0$
  for $t = 1, \dots, T$ **do**
    $\boldsymbol{p}_{t+1/2} \coloneqq \boldsymbol{p}_t + \frac{\Delta t}{2}\widehat{\mathscr{G}}_t(\theta)$
    $\boldsymbol{\theta}_{t+1/2} \coloneqq \boldsymbol{\theta}_t + \frac{\Delta t}{2}\boldsymbol{p}_t$
    $\boldsymbol{\xi}_{t+1/2} \coloneqq \boldsymbol{\xi}_t + \frac{\Delta t}{2}(\boldsymbol{p}_{t+1/2}\odot\boldsymbol{p}_{t+1/2} - 1)\mu^{-1}$
    $\boldsymbol{r}_{t+1/2} \coloneqq \exp(-\lambda\Delta t/2)(\boldsymbol{r}_t - \boldsymbol{p}_{t+1/2}\lambda^{-1}) + \boldsymbol{p}_{t+1/2}\lambda^{-1}$
    $\hat{\boldsymbol{p}}_{t+1/2} \coloneqq \exp(-\Delta t\boldsymbol{\xi})\boldsymbol{p}_{t+1/2} + (1-\exp(-2\Delta t\boldsymbol{\xi}))\,\widehat{\boldsymbol{\sigma}}^2_{(t)}\boldsymbol{r}_{t+1/2}$
    $\boldsymbol{r}_{t+1} \coloneqq \exp(-\lambda\Delta t/2)(\boldsymbol{r}_{t+1/2} - \hat{\boldsymbol{p}}_{t+1/2}\lambda^{-1}) + \hat{\boldsymbol{p}}_{t+1/2}\lambda^{-1}$
    $\boldsymbol{\xi}_{t+1} \coloneqq \boldsymbol{\xi}_{t+1/2} + \frac{\Delta t}{2}(\hat{\boldsymbol{p}}_{t+1/2}\odot\hat{\boldsymbol{p}}_{t+1/2} - 1)\mu^{-1}$
    $\boldsymbol{\theta}_{t+1} \coloneqq \boldsymbol{\theta}_{t+1/2} + \frac{\Delta t}{2}\hat{\boldsymbol{p}}_{t+1/2}$
    $\boldsymbol{p}_{t+1} \coloneqq \hat{\boldsymbol{p}}_{t+1/2} + \frac{\Delta t}{2}\widehat{\mathscr{G}}_t(\theta)$
    update $\mathcal{I}_t$ according to the replacement rule
  end for
  Return $(\boldsymbol{\theta}_t)_{1\le t\le T}$

[1] Note that the additional gradient noise $\boldsymbol{\eta}$ breaks the symmetry of the splitting scheme

	Reference	CCAdGLD	CCAdLD
IAT of $\boldsymbol{\theta}_1$	26.10	65.1030	51.0352
IAT of $\boldsymbol{\theta}_1^2$	33.9072	68.1341	56.9623

Table 5.1: Integrated autocorrelation times for logistic regression on MNIST

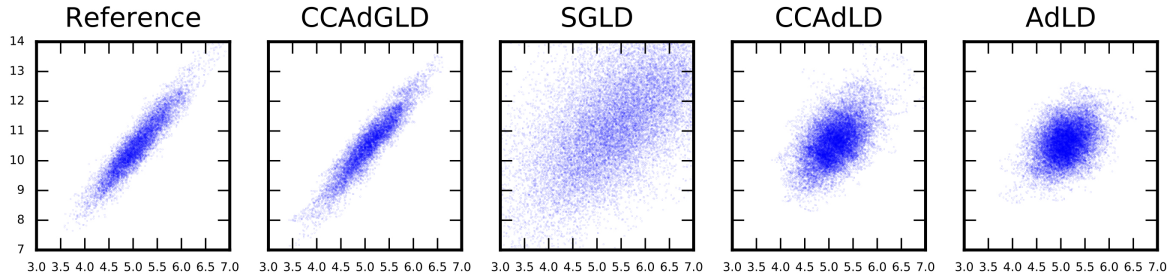

Figure 5.1: Traces of samples of various methods targeting the posterior of the 2 dimensional logistic regression model.

5.3 Numerical experiments

In this section we assess the sampling properties of the following methods: (i) SGLD: The Stochastic Gradient Langevin Dynamics method, (ii) AdLD: An adaptive form Langevin dynamics as discussed in [58, 27, 111] which automatically adjusts a Langevin friction parameter, but does not use a memory kernel, (iii) CCAdLD: Covariance controlled adaptive Langevin dynamics as proposed in [111], and (iv) CCAdGLD: the method introduced in this chapter.

All reference solutions were computed with a BAOAB-Langevin integrator using exact gradients. Multiple runs with varying stepsizes were performed and the effective samples size monitored in order to make sure that both sampling error and discretization bias were reduced to a sufficient small order.

5.3.1 Two-dimensional test cases

In order to flag up performance and possible shortcomings of the methods proposed here in a controlled setup, we first consider sampling of a two dimensional dimensional Laplace target density $\boldsymbol{p}(\boldsymbol{\theta}) \propto \exp(-|\boldsymbol{\theta}|_1), \boldsymbol{\theta} \in \mathbb{R}^2$ where we perturb the gradient $\mathscr{G}(\boldsymbol{\theta})$ by a synthetically generated Gaussian noise process with auto-covariance function $\boldsymbol{K}(t) = \boldsymbol{I}_2 e^{-t/2}$. A harmonic weighting is used in the case of CCAdGLD and CCAdLD and covariance estimates are additionally perturbed by white noise of variance $1/2$. Figure 5.2 shows that the CCAdGLD method clearly outperforms all other methods in terms of accuracy. The sample trace of CCAdGLD can not be visually distinguished from the reference solution. Other gradient methods yield samples of a vastly increased variance in comparison to the reference solution. For all schemes a stepsize of $\Delta t = 10^{-1}$ was used. Table 5.2 confirms that a comparison based on this parametrisation is fair since the integrated autocorrelation time (IAT) for all methods is of a similar order.

In order to test the robustness of the methods in situations where the assumption of Gaussianity and stationarity of the gradient noise is not expected to hold, we consider

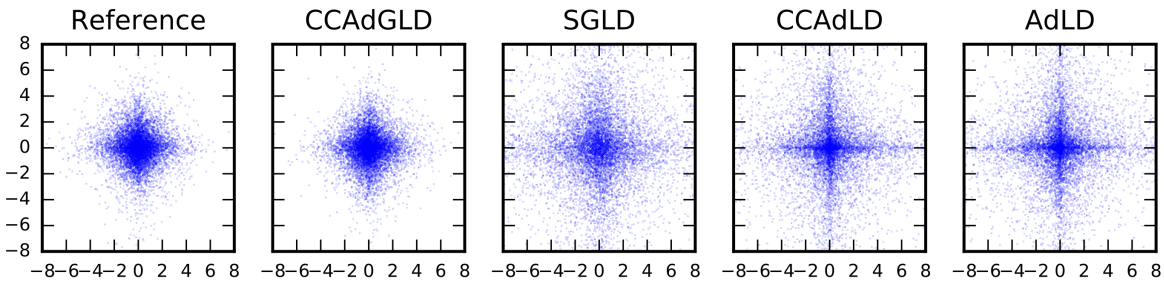

Figure 5.2: Traces of samples of various methods targeting the two dimensional dimensional Laplace density.

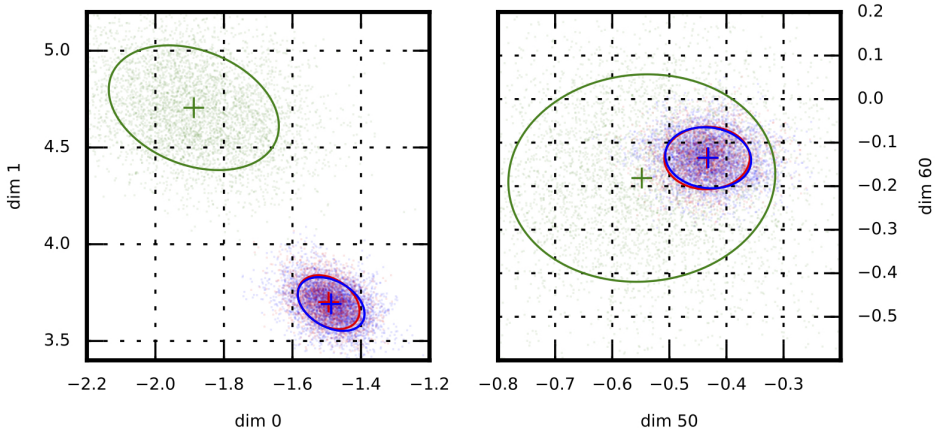

Figure 5.3: Cross-sections in randomly selected pairs of dimensions of the sample trajectories of CCAdGLD (blue), CCAdLD (green), and the reference sample (red). Ellipsoids show the empirical 95 percent credibility regions calculated from the empirical covariance and mean of the respective samples.

a logistic regression problem in $\mathbb{R}^2$ using the likelihood function

$$\pi(\{\boldsymbol{x}_i, y_i\}_{i=1}^N|\boldsymbol{\beta}) \propto \prod_{i=1}^N 1/\left(1+\exp(-y_i\boldsymbol{\beta}^\top \boldsymbol{x}_i)\right),$$

and a non-informative Gaussian prior

$$\pi(\boldsymbol{\beta}) \propto \exp(-\boldsymbol{\beta}^\top\boldsymbol{\beta}/(2\sigma^2)), \sigma^2 = 500.$$

We use synthetic data, which we generate by first sampling N = 1000 data points $(\mathbf{x}_i)_{i=1}^1 000$ from an isotropic Gaussian of unit variance and then assign class labels $y_i \in \{0,1\}$ with probability $\pi(y_i|\boldsymbol{x}_i, \boldsymbol{\beta} = (5,10))$. We use minbatches of size 50 to compute the gradient estimate $\hat{\mathscr{G}}_t$ and refresh datapoints according to the k-RR scheme described in section 5.1.2 with an overlay of $k = 5$ data points. Similarly as in the previous example, we observe that the reference sample and the sample generated by CCAdGLD are visually indistinguishable (Figure 5.1), whereas samples of other methods do vastly deviate from the reference sample.

	CCAdGLD	SGLD	CCAdLD	AdLD
IAT	84.23	126.38	87.64	136.12
Mean	0.06 ± 0.06	-0.03 ± 0.19	0.07 ± 0.14	0.01 ± 0.19
Var	1.95 ± 0.19	15.56 ± 2.20	11.83 ± 1.36	13.41 ± 2.20

Table 5.2: IAT and first moments for θ_1 for sampling of the Laplace target density

5.3.2 Logistic regression on MNIST

We consider the MNIST classification problem as presented in [111]; we extract all instances of the digits "7" and "9" from the full dataset, and train a Bayesian logistic regression model to classify instances of these digits. We preprocess the data by subsampling each image to size 10×10, using SciPy's in-built spline interpolation library. Training is then run on $12,251$ images. We train using CCAdGLD, CCAdLD. Minibatches of size 500 are selected using the Poisson replacement scheme of Section 5.1.2, with $p = 0.05$. Figure 5.3 shows that the CCAdGLD methods recovers the reference sample accurately, while CCAdLD deviates strongly both in terms of the covariance structure as well as the estimated mean. We expect the deviation in the mean to be due to entropic effects caused by the failure of CCAdLD to recover the correct variance structure. Again, the comparison between the methods is fair since the integrated auto-correlation time of CCAdGLD and CCAdLD is of the same order as Table 5.1 shows.

Chapter 6

Conclusion

In this thesis we discussed asymptotic properties and the numerical discretisation of instances of the generalised Langevin equation which can be represented in the Markovian forms (3.1-3.4) and (3.22). We extended previous results on the geometric ergodicity of the solution processes of such quasi-Markovian generalised Langevin equations in Section 3.4. Although the ergodicity results for GLEs representable in the form (3.1-3.4) are straightforward extensions of already existing results on the ergodicity properties of the underdamped Langevin equation or variants of (3.1-3.4) (see e.g. [81, 12, 96]), we believe that by deriving sufficient criteria for geometric ergodicity in the form of algebraic conditions on the matrices $\mathbf{\Gamma}, \mathbf{\Sigma}$ in the Markovian representation of the GLE results which are of practical relevance. In particular, we believe that such results are helpful for practitioners without a background in stochastic analysis. The ergodicity result (Theorem 3.4.4) on Markovian reformulations of the form (3.22) is to the best of our knowledge novel and we believe it (or extensions of this theorem) to be relevant in a wide range of modelling applications where the assumption of the random force being independent of the configurational variable $\boldsymbol{q}$ is not valid. In [108] we showed in the case of the underdamped Langevin equation that previous ergodicity results can be extended in a straightforward way to variants of the Langevin equation corresponding to non-equilibrium modelling situations where for example the deterministic force is not given as the gradient of a potential function and the fluctuation-dissipation relation between between stochastic and dissipative forces is relaxed. We expect that the ergodicity results on (3.22) as well as the ergodicity results on (3.1-3.4) can be extended in a similar way.

We proposed new splitting schemes for the Markovian representations of quasi-Markovian generalised Langevin equations based on a decomposition of the corresponding vector field into a Hamiltonian part and the remaining stochstic linear part. The schemes proposed in Section 4.1 can be seen as adaptations of schemes previously proposed for the underdamped Langevin equation and are particularly well suited for GLEs driven by a stationary noise process without cross-correlation terms. In numerical experiments we find that among the proposed schemes the one corresponding to the integration sequence B→A→O→A→B performs best in terms of the observed error due to discretisation in ergodic averages. We support theses observations

- with an analysis of the discretisation error in the situation where the potential function is harmonic.
- with a formal analysis of the discretisation error in the overdamped limit of the GLE.

Moreover, we found that when used in combination with the pre-optimised memory kernels as provided by [21], the BAOAB integration scheme leads to improved mixing properties and stability for large stepsizes in comparison to sampling schemes based on an underdamped Langevin equation with tuned friction coefficient. These results motivate to more closely inspect the decay properties (e.g. within the framework of [28]) of the semigroup operators associated with instances of (3.1-3.4) for which the parametrisation of the matrices $\mathbf{\Gamma}, \mathbf{\Sigma}$ resembles the structure of the matrices derived within the framework of [21]. A good starting point for such an analysis would be a generalised Langevin equation with a memory kernel as considered in [126] (see also Section 3.1.3).

We showed in Section 3.5 that the H-OU type splitting schemes for the GLE which we discussed in Section 4.1 behave consistently in the white noise and overdamped limit. Moreover, it follows from the analysis presented there that the stability of the numerical schemes is not affected in the respective asymptotic limits. We therefore provide a class of numerical integration schemes which can equally be applied to the overdamped, the underdamped and the generalised Langevin equation. Moreover, we rigorously proved geometric ergodicity in Section 4.4 for the numerical schemes proposed in Section 4.1 under the assumption that the domain of the configurational variable is bounded, which implies that a central limit theorem holds for observables contained in the respective Banach spaces. We therefore provide the theoretical foundation which justifies the usage of GLE based dynamics for sampling purposes. We also provide instructions for the metropolisation of GLE schemes in Section 4.3. For the reasons explained in this section the usefulness of such schemes in practical applications is, however, questionable.

In Section 4.7 we proposed new splitting methods which incorporate a modified O-step allowing efficient numerical integration of GLEs with sparse complex memory kernels and non-stationary random forces. These methods rely on a further splitting of the OU equation at each timestep and can be combined with multiple timestepping in order to increase the accuracy of the approximate solution of the O-step. The decomposition of the O-step is chosen such that matrix-free operations can be used in the computation of the solution of each substep. While we provide analytically derived results on e.g. the parameter dependent numerical stability of such methods, a detailed evaluation of the performance of the numerical methods proposed in Section 4.7 in the form of systematic numerical simulations is still missing. Apart from the AR-p splitting method, the numerical methods which we proposed in this chapter can be also applied to instances of the underdamped Langevin equations. In [108] we demonstrated that the FD-p splitting scheme in combinations with multiple timestepping allows efficient integration of the underdamped Langevin equation in non-equilibrium models where a standard BAOAB integration scheme would lead to excessively high computational costs per timestep. Another interesting potential application of the integration schemes proposed Section 4.7 is dissipative particle dynamics (DPD). Numerical integrators for DPD typically rely on a Shardlow subsplitting [112] of the O-step which results in an algorithm which is difficult to parallelise. At this point it remains open how the methods of Section 4.7 compare to existing integration schemes for DPD.

The GLE based adaptive thermostat CCadGLD method which we propose in Chapter 5 is to the best of our knowledge the first attempt in extending the setup of stochastic gradient methods from a Markovian to a non-Markovian framework. As such the formulation of CCadGLD is novel. We demonstrate in numerical experiments that by explicitly incorporating memory effects in the thermostat dynamics via the GLE that

we can accurately dissipate time correlated gradient noise if the auto-correlation function of the noise is known a priori. In addition we derive the form of the autocorrelation function of the gradient noise which arises when minibatches are only partially refreshed according to certain replacement rules between Markov chain time steps in Bayesian posterior sampling.

Bibliography

[1] A. Abdulle, G. Vilmart, and K. C. Zygalakis. Long time accuracy of Lie-Trotter splitting methods for Langevin dynamics. *SIAM Journal on Numerical Analysis*, 53(1):1–16, 2015.

[2] S. Adelman and J. Doll. Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *The Journal of chemical physics*, 64(6):2375–2388, 1976.

[3] S. Ahn, A. Korattikara, N. Liu, S. Rajan, and M. Welling. Large-scale distributed Bayesian matrix factorization using stochastic gradient MCMC. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 9–18. ACM, 2015.

[4] S. Ahn, B. Shahbaba, and M. Welling. Distributed stochastic gradient MCMC. In *International conference on machine learning*, pages 1044–1052, 2014.

[5] E. Akhmatskaya, N. Bou-Rabee, and S. Reich. A comparison of generalized hybrid Monte Carlo methods with and without momentum flip. *Journal of Computational Physics*, 228(6):2256–2265, 2009.

[6] C. Andrieu, É. Moulines, et al. On the ergodicity properties of some adaptive MCMC algorithms. *The Annals of Applied Probability*, 16(3):1462–1505, 2006.

[7] A. D. Baczewski and S. D. Bond. Numerical integration of the extended variable generalized Langevin equation with a positive Prony representable memory kernel. *The Journal of chemical physics*, 139(4):044107, 2013.

[8] A. K. Balan, V. Rathod, K. P. Murphy, and M. Welling. Bayesian dark knowledge. In *Advances in Neural Information Processing Systems*, pages 3438–3446, 2015.

[9] D. Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[10] H. Bauer. *Measure and integration theory*, volume 26. Walter de Gruyter, 2001.

[11] H. Bauer and R. B. Burckel. *Probability theory and elements of measure theory*. Academic Press London, 1981.

[12] L. R. Bellet. Ergodic properties of Markov processes. In *Open quantum systems II*, pages 1–39. Springer, 2006.

[13] G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near-to-the identity symplectic mappings with application to symplectic integration algorithms. *Journal of Statistical Physics*, 74(5):1117–1143, 1994.

[14] R. Bhatia and P. Rosenthal. How and why to solve the operator equation AX-XB= Y. *Bulletin of the London Mathematical Society*, 29(1):1–21, 1997.

[15] R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 60(2):185–201, 1982.

[16] C. M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

[17] N. Bou-Rabee. On Metropolis integrators for molecular dynamics. *arXiv preprint arXiv:1309.5007*, 2013.

[18] N. Bou-Rabee and H. Owhadi. Long-run accuracy of variational integrators in the stochastic context. *SIAM Journal on Numerical Analysis*, 48(1):278–297, 2010.

[19] P. Carmona. Existence and uniqueness of an invariant measure for a chain of oscillators in contact with two heat baths. *Stochastic Processes and their Applications*, 117(8):1076–1092, 2007.

[20] M. Ceriotti. Gle4md: http://gle4md.org.

[21] M. Ceriotti, G. Bussi, and M. Parrinello. Langevin equation with colored noise for constant-temperature molecular dynamics simulations. *Physical review letters*, 102(2):020601, 2009.

[22] M. Ceriotti, G. Bussi, and M. Parrinello. Colored-noise thermostats à la carte. *Journal of Chemical Theory and Computation*, 6(4):1170–1180, 2010.

[23] T. Chen, E. Fox, and C. Guestrin. Stochastic gradient Hamiltonian Monte Carlo. In *International Conference on Machine Learning*, pages 1683–1691, 2014.

[24] N. Chopin, T. Lelièvre, and G. Stoltz. Free energy methods for Bayesian inference: efficient exploration of univariate Gaussian mixture posteriors. *Statistics and Computing*, 22(4):897–916, 2012.

[25] G. Ciccotti and J.-P. Ryckaert. On the derivation of the generalized Langevin equation for interacting Brownian particles. *Journal of Statistical Physics*, 26(1):73–82, 1981.

[26] E. Darve, J. Solomon, and A. Kia. Computing generalized Langevin equations and generalized Fokker–Planck equations. *Proceedings of the National Academy of Sciences*, 106(27):10884–10889, 2009.

[27] N. Ding, Y. Fang, R. Babbush, C. Chen, R. D. Skeel, and H. Neven. Bayesian sampling using stochastic gradient thermostats. In *Advances in neural information processing systems*, pages 3203–3211, 2014.

[28] J. Dolbeault, C. Mouhot, and C. Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Transactions of the American Mathematical Society*, 367(6):3807–3828, 2015.

[29] J. Doll and D. Dion. Generalized Langevin equation approach for atom/solid–surface scattering: Numerical techniques for Gaussian generalized Langevin dynamics. *The Journal of Chemical Physics*, 65(9):3762–3766, 1976.

[30] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222, 1987.

[31] A. Duncan, G. Pavliotis, and K. Zygalakis. Nonreversible Langevin samplers: Splitting schemes, analysis and implementation. *arXiv preprint arXiv:1701.04247*, 2017.

[32] H. Dym and H. P. McKean. *Gaussian processes, function theory, and the inverse spectral problem*. Courier Corporation, 2008.

[33] J.-P. Eckmann and M. Hairer. Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators. *Communications in Mathematical Physics*, 212(1):105–164, 2000.

[34] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Entropy production in nonlinear, thermally driven Hamiltonian systems. *Journal of statistical physics*, 95(1):305–331, 1999.

[35] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Communications in Mathematical Physics*, 201(3):657–697, 1999.

[36] R. D. Engle, R. D. Skeel, and M. Drees. Monitoring energy drift with shadow Hamiltonians. *Journal of Computational Physics*, 206(2):432–452, 2005.

[37] G. Ford, M. Kac, and P. Mazur. Statistical mechanics of assemblies of coupled oscillators. *Journal of Mathematical Physics*, 6(4):504–515, 1965.

[38] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Academic press, 2001.

[39] C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer-Verlag, 1994.

[40] D. Givon, R. Kupferman, and O. H. Hald. Existence proof for orthogonal dynamics and the Mori-Zwanzig formalism. *Israel Journal of Mathematics*, 145(1):221–241, 2005.

[41] D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17(6):R55, 2004.

[42] H. Goldstein. *Classical mechanics*. Pearson Education India, 2011.

[43] E. Hairer and C. Lubich. The life-span of backward error analysis for numerical integrators. *Numerische Mathematik*, 76(4):441–462, 1997.

[44] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, volume 31. Springer Science & Business Media, 2006.

[45] M. Hairer and J. C. Mattingly. Yet another look at Harris ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63, pages 109–117. Springer, 2011.

[46] P. Hänggi. Generalized Langevin equations: A useful tool for the perplexed modeller of nonequilibrium fluctuations? *Stochastic dynamics*, pages 15–22, 1997.

[47] T. E. Harris. The existence of stationary measures for certain Markov processes. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 113–124, 1956.

[48] L. Hörmander. The analysis of linear partial differential operators. III, volume 274 of Grundlehren der Mathematischen Wissenschaften [fundamental principles of mathematical sciences], 1985.

[49] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 1990.

[50] J. K. Hunter and B. Nachtergaele. *Applied analysis*. World Scientific Publishing Co Inc, 2001.

[51] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland mathematical library. North-Holland Publishing Company, 1981.

[52] A. J. Izenman and C. J. Sommer. Philatelic mixtures and multimodal densities. *Journal of the American Statistical association*, 83(404):941–953, 1988.

[53] V. Jakišć and C.-A. Pillet. Ergodic properties of the non-Markovian Langevin equation. *Letters in Mathematical Physics*, 41(1):49–57, 1997.

[54] V. Jakšić and C.-A. Pillet. Spectral theory of thermal relaxation. *Journal of Mathematical Physics*, 38(4):1757–1780, 1997.

[55] V. Jakšić and C.-A. Pillet. Ergodic properties of classical dissipative systems i. *Acta mathematica*, 181(2):245–282, 1998.

[56] A. Jasra, C. C. Holmes, and D. A. Stephens. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, pages 50–67, 2005.

[57] O. G. Jepps, G. Ayton, and D. J. Evans. Microscopic expressions for the thermodynamic temperature. *Physical Review E*, 62(4):4757, 2000.

[58] A. Jones and B. Leimkuhler. Adaptive stochastic methods for sampling driven molecular systems. *The Journal of chemical physics*, 135(8):084125, 2011.

[59] R. Joubaud, G. Pavliotis, and G. Stoltz. Langevin dynamics with space-time periodic nonequilibrium forcing. *Journal of Statistical Physics*, 158(1):1–36, 2015.

[60] L. Kantorovich. Generalized Langevin equation for solids. I. Rigorous derivation and main properties. *Physical Review B*, 78(9):094304, 2008.

[61] W. Kliemann. Recurrence and invariant measures for degenerate diffusions. *The annals of probability*, pages 690–707, 1987.

[62] P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2011.

[63] M. Kopec. Weak backward error analysis for overdamped Langevin processes. *IMA Journal of Numerical Analysis*, 35(2):583–614, 2014.

[64] M. Kopec. Weak backward error analysis for Langevin process. *BIT Numerical Mathematics*, 55(4):1057–1103, 2015.

[65] R. Kupferman. Fractional kinetics in Kac–Zwanzig heat bath models. *Journal of statistical physics*, 114(1):291–326, 2004.

[66] R. Kupferman, A. Stuart, J. Terry, and P. Tupper. Long-term behaviour of large mechanical systems with random initial data. *Stochastics and Dynamics*, 2(04):533–562, 2002.

[67] L. D. Landau, E. Lifshitz, and L. Pitaevskii. Statistical physics (course of theoretical physics, volume 5). *3rd. Edition*, 1984.

[68] Y. LeCun and C. Cortes. MNIST handwritten digit database. 2010.

[69] H. Lei, N. A. Baker, and X. Li. Data-driven parameterization of the generalized Langevin equation. *Proceedings of the National Academy of Sciences*, 113(50):14183–14188, 2016.

[70] B. Leimkuhler and C. Matthews. Robust and efficient configurational molecular sampling via Langevin dynamics. *The Journal of chemical physics*, 138(17):05B601_1, 2013.

[71] B. Leimkuhler and C. Matthews. *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods*. Interdisciplinary Applied Mathematics. Springer, 5 2015.

[72] B. Leimkuhler, C. Matthews, and G. Stoltz. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA Journal of Numerical Analysis*, 36(1):13–79, 2015.

[73] B. Leimkuhler and S. Reich. Simulating Hamiltonian dynamics, volume 14 of Cambridge monographs on applied and computational mathematics. *Cambridge University Press, Cambridge*, 2:18, 2004.

[74] B. Leimkuhler and X. Shang. Adaptive thermostats for noisy gradient systems. *SIAM Journal on Scientific Computing*, 38(2):A712–A736, 2016.

[75] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations: A mathematical perspective*. World Scientific, 2010.

[76] T. Lelièvre and G. Stoltz. Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica*, 25:681–880, 2016.

[77] C. Li, C. Chen, D. E. Carlson, and L. Carin. Preconditioned stochastic gradient Langevin dynamics for deep neural networks. In *AAAI*, volume 2, page 4, 2016.

[78] Z. Li, X. Bian, X. Li, and G. E. Karniadakis. Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism. *The Journal of chemical physics*, 143(24):243128, 2015.

[79] Z. Li, H. S. Lee, E. Darve, and G. E. Karniadakis. Computing the non-Markovian coarse-grained interactions derived from the Mori–Zwanzig formalism in molecular systems: Application to polymer meltswanzig formalism in molecular systems: Application to polymer melts. *The Journal of chemical physics*, 146(1):014104, 2017.

[80] S. H. Lim and J. Wehr. Homogenization of a class of non-Markovian langevin equations with an application to thermophoresis. *arXiv preprint arXiv:1704.00134*, 2017.

[81] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.

[82] J. C. Mattingly, A. M. Stuart, and M. V. Tretyakov. Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM Journal on Numerical Analysis*, 48(2):552–577, 2010.

[83] S. Melchionna. Design of quasisymplectic propagators for Langevin dynamics. *The Journal of chemical physics*, 127(4):044108, 2007.

[84] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes I: Criteria for discrete-time chains. *Advances in Applied Probability*, 24(3):542–574, 1992.

[85] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes II: Continuous-time processes and sampled chains. *Advances in Applied Probability*, 25(3):487–517, 1993.

[86] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[87] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics*. Springer Science & Business Media, 2013.

[88] H. Mori. A continued-fraction representation of the time-correlation functions. *Progress of Theoretical Physics*, 34(3):399–416, 1965.

[89] G. P. Morriss and D. J. Evans. *Statistical Mechanics of Nonequilbrium Liquids*. ANU Press, 2013.

[90] J. A. Morrone, T. E. Markland, M. Ceriotti, and B. Berne. Efficient multiple time scale molecular dynamics: Using colored noise thermostats to stabilize resonances. *The Journal of chemical physics*, 134(1):014103, 2011.

[91] K. P. Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[92] R. M. Neal. *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media, 2012.

[93] H. Ness, L. Stella, C. Lorenz, and L. Kantorovich. Applications of the generalized Langevin equation: Towards a realistic description of the baths. *Physical Review B*, 91(1):014301, 2015.

[94] H. Ness, L. Stella, C. Lorenz, and L. Kantorovich. Temperature and length dependence of the nonequilibrium heat transport in atomic chains between two realistic thermal baths: a generalised Langevin equation approach. *arXiv preprint arXiv:1612.00990*, 2016.

[95] B. Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.

[96] M. Ottobre and G. Pavliotis. Asymptotic analysis for the generalized Langevin equation. *Nonlinearity*, 24(5):1629, 2011.

[97] S. Patterson and Y. W. Teh. Stochastic gradient Riemannian Langevin dynamics on the probability simplex. In *Advances in Neural Information Processing Systems*, pages 3102–3110, 2013.

[98] G. A. Pavliotis. *Stochastic processes and applications*. Springer, 2016.

[99] S. Redon, G. Stoltz, and Z. Trstanova. Error analysis of modified Langevin dynamics. *Journal of Statistical Physics*, 164(4):735–771, 2016.

[100] L. Rey-Bellet. Statistical mechanics of anharmonic lattices. *arXiv preprint math-ph/0303021*, 2003.

[101] L. Rey-Bellet. Open classical systems. In *Open Quantum Systems II*, pages 41–78. Springer, 2006.

[102] L. Rey-Bellet and L. E. Thomas. Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Communications in mathematical physics*, 225(2):305–329, 2002.

[103] S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)*, 59(4):731–792, 1997.

[104] J. Rosenthal and G. Roberts. Coupling and ergodicity of adaptive MCMC. *Journal of Applied Probablity*, 44:458–475, 2007.

[105] W. Rudin. Functional analysis. International series in pure and applied mathematics, 1991.

[106] W. Rudin. *Fourier analysis on groups*. Courier Dover Publications, 2017.

[107] H. H. Rugh. Dynamical approach to temperature. *Physical review letters*, 78(5):772, 1997.

[108] M. Sachs, B. Leimkuhler, and V. Danos. Langevin dynamics with variable coefficients and nonconservative forces: From stationary states to numerical methods. *Entropy*, 19(12):647, 2017.

[109] Z. Schuss. Diffusion and stochastic processes: an analytical approach. *Springer series on Applied mathematical sciences*, 170, 2010.

[110] C. Schwab and C. J. Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, 20:291–467, 2011.

[111] X. Shang, Z. Zhu, B. Leimkuhler, and A. J. Storkey. Covariance-controlled adaptive Langevin thermostat for large-scale Bayesian sampling. In *Advances in Neural Information Processing Systems*, pages 37–45, 2015.

[112] T. Shardlow. Splitting for dissipative particle dynamics. *SIAM Journal on Scientific computing*, 24(4):1267–1282, 2003.

[113] A. Sokal. Monte Carlo methods in statistical mechanics: foundations and new algorithms. In *Functional integration*, pages 131–192. Springer, 1997.

[114] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

[115] L. Stella, C. Lorenz, and L. Kantorovich. Generalized Langevin equation: An efficient approach to nonequilibrium molecular dynamics of open systems. *Physical Review B*, 89(13):134303, 2014.

[116] D. W. Stroock and S. R. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, volume 3, pages 333–359, 1972.

[117] M. Suzuki. Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations. *Physics Letters A*, 146(6):319–323, 1990.

[118] D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002.

[119] Y.-F. Tang and Y.-H. Long. Formal energy of symplectic scheme for Hamiltonian systems and its applications (II). *Computers & mathematics with applications*, 27(12):31–39, 1994.

[120] Y. W. Teh, A. H. Thiery, and S. J. Vollmer. Consistency and fluctuations for stochastic gradient Langevin dynamics. *Journal of Machine Learning Research*, 17:1–33, 2016.

[121] R. C. Tolman. *The principles of statistical mechanics*. Courier Corporation, 1938.

[122] L. Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review*, 159(1):98, 1967.

[123] C. Villani. *Hypocoercivity*. American Mathematical Soc.

[124] S. J. Vollmer, K. C. Zygalakis, et al. (non-) asymptotic properties of stochastic gradient Langevin dynamics. *arXiv preprint arXiv:1501.00438*, 2015.

[125] M. Welling and Y. W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 681–688, 2011.

[126] X. Wu, B. R. Brooks, and E. Vanden-Eijnden. Self-guided Langevin dynamics via generalized Langevin equation. *Journal of computational chemistry*, 37(6):595–601, 2016.

[127] H. Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150(5-7):262–268, 1990.

[128] F. Zhang. *The Schur complement and its applications*, volume 4. Springer Science & Business Media, 2006.

[129] K. Zhou, J. C. Doyle, K. Glover, et al. *Robust and optimal control*, volume 40. Prentice hall New Jersey, 1996.

[130] R. Zwanzig. Memory effects in irreversible thermodynamics. *Physical Review*, 124(4):983, 1961.

[131] R. Zwanzig. Nonlinear generalized Langevin equations. *Journal of Statistical Physics*, 9(3):215–220, 1973.

Appendix A

Auxiliary material on linear algebra

The following Lemma A.0.1 is repeatedly used in the proofs of Lemma 3.1.1 and Lemma 3.4.3, as well as in Example 3.4.1 to show the positive (semi-)definiteness of symmetric matrices.

Lemma A.0.1. *Let A be a symmetric block structured matrix of the form*

$$\boldsymbol{A} := \begin{pmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} \\ \boldsymbol{A}_{1,2}^T & \boldsymbol{A}_{2,2} \end{pmatrix} \in \mathbb{R}^{n+m \times n+m}$$

(i) If $\boldsymbol{A}_{2,2}$ is positive definite, then $\boldsymbol{A}$ is positive (semi-)definite if and only if

$$\boldsymbol{A}_{1,1} - \boldsymbol{A}_{1,2}\boldsymbol{A}_{2,2}^{-1}\boldsymbol{A}_{1,2}^T$$

is positive (semi-)definite

(ii) If $\boldsymbol{A}_{1,1}$ is positive definite, then $\boldsymbol{A}$ is positive (semi-)definite if and only if

$$\boldsymbol{A}_{2,2} - \boldsymbol{A}_{1,2}^T\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}$$

is positive (semi-)definite

(iii) Let $\boldsymbol{A}_{2,2}^g$ denote a generalised inverse of $\boldsymbol{A}_{2,2}$, i.e., $\boldsymbol{A}_{2,2}^g$ is a $m \times m$ matrix which satisfies

$$\boldsymbol{A}_{2,2}\boldsymbol{A}_{2,2}^g\boldsymbol{A}_{2,2} = \boldsymbol{A}_{2,2}.$$

The matrix $\boldsymbol{A}$ is positive semi-definite if and only if the matrices $\boldsymbol{A}_{2,2}$ and $\boldsymbol{A}_{1,1} - \boldsymbol{A}_{1,2}\boldsymbol{A}_{2,2}^g\boldsymbol{A}_{1,2}^T$ are positive semi-definite, and

$$(\boldsymbol{I} - \boldsymbol{A}_{2,2}\boldsymbol{A}_{2,2}^g)\boldsymbol{A}_{1,2}^T = \boldsymbol{0},$$

i.e., the span of the column vectors of $\boldsymbol{A}_{1,2}$ is contained in the span of the column vectors of $\boldsymbol{A}_{1,1}$.

Proof. The statements (i) and (ii) follow from Theorem 1.12 in [128]. Statement (iii) corresponds to Theorem 1.20 in the same reference. □

The following Lemma is used in Section 3.1.1 to derive a uniqueness result on the form of the invariant measure for a given parametrisation of the SDE (3.1-3.3).

Lemma A.0.2. *Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^n$. There exists a unique solution for the Sylvester equation*

$$\boldsymbol{AX} + \boldsymbol{XB} = \boldsymbol{C},$$

if an only if $\boldsymbol{A}$ and $\boldsymbol{B}$, have no common eigenvalues.

Proof. See [14]. □

Appendix B

Algorithms for the GLE and adaptive (generalised) Langevin dynnamics

The following is a list of algorithms considered in numerical simulations presented in this thesis; see Section 4.8. The algorithms (B.1)-(B.4) are constructed as H-OU splitting schemes. Details on their construction can be found in Section 4.1. The algorithms (B.5) and (B.6) were proposed in [7]. We provide further details and a derivation of the limit-method of (B.6) in Section 4.1. We present the scheme [115] proposed in (B.7) in an condensed form, i.e., algorithm (B.7) is the resulting updating when the method proposed in [115] is applied to a GLE with scalar valued memory kernel function $\boldsymbol{K}(t) = ce^{-\frac{t}{\tau}}\cos(\alpha t)$.

gle-BAOAB

$$\begin{aligned}
\boldsymbol{p}^{k+1/2} &= \boldsymbol{p}^k - \frac{\Delta t}{2}\nabla_{\boldsymbol{q}} U(\boldsymbol{q}^k) \\
\boldsymbol{q}^{k+1/2} &= \boldsymbol{q}^k + \frac{\Delta t}{2}\boldsymbol{q}^{k+1/2} \\
\begin{pmatrix} \hat{\boldsymbol{p}}^{k+1/2} \\ \boldsymbol{s}^{k+1} \end{pmatrix} &= \boldsymbol{F}_{\Delta t}\begin{pmatrix} \boldsymbol{p}^{k+1/2} \\ \boldsymbol{s}^{k} \end{pmatrix} + \boldsymbol{S}_{\Delta t \Delta t}\mathscr{R}^k \\
\boldsymbol{q}^{k+1} &= \boldsymbol{q}^{k+1/2} + \frac{\Delta t}{2}\hat{\boldsymbol{p}}^{k+1/2} \\
\boldsymbol{p}^{k+1} &= \hat{\boldsymbol{p}}^{k+1/2} - \frac{\Delta t}{2}\nabla_{\boldsymbol{q}} U(\boldsymbol{q}^{k+1})
\end{aligned} \tag{B.1}$$

gle-OBABO, Ceriotti's method

$$\begin{aligned}
\begin{pmatrix} \hat{\boldsymbol{p}}^{k} \\ \hat{\boldsymbol{s}}_i^{k} \end{pmatrix} &= \boldsymbol{F}_{\Delta t}\begin{pmatrix} \boldsymbol{p}^{k} \\ \boldsymbol{s}^{k} \end{pmatrix} + \boldsymbol{S}_{\Delta t \Delta t}\mathscr{R}^k \\
\boldsymbol{p}^{k+1/2} &= \hat{\boldsymbol{p}}^{k} - \frac{\Delta t}{2}\nabla_{\boldsymbol{q}} U(\boldsymbol{q}^{k+1}) \\
\boldsymbol{q}^{k+1} &= \boldsymbol{q}^{k} + \Delta t \boldsymbol{p}^{k+1/2} \\
\tilde{\boldsymbol{p}}^{k+1} &= \hat{\boldsymbol{p}}^{k+1/2} - \frac{\Delta t}{2}\nabla_{\boldsymbol{q}} U(\boldsymbol{q}^{k+1}) \\
\begin{pmatrix} \boldsymbol{p}^{k+1} \\ \boldsymbol{s}^{k+1} \end{pmatrix} &= \boldsymbol{F}_{\Delta t \Delta t}\begin{pmatrix} \hat{\boldsymbol{p}}^{k+1} \\ \hat{\boldsymbol{s}}_i^{k} \end{pmatrix} + \boldsymbol{S}_{\Delta t \Delta t}\hat{\mathscr{R}}^k
\end{aligned} \tag{B.2}$$

gle-SAS

$$\begin{aligned}
\begin{pmatrix}\hat{\boldsymbol{p}}^k\\ \hat{\boldsymbol{s}}_i^k\end{pmatrix} &= -\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}_{m+n}-\boldsymbol{F}_{\Delta t/2})\nabla_{\boldsymbol{q}}U(\boldsymbol{q}^k)+\boldsymbol{F}_{\Delta t/2}\begin{pmatrix}\boldsymbol{p}^k\\ \boldsymbol{s}^k\end{pmatrix}+\boldsymbol{S}_{\Delta t/2}\mathscr{R}^k\\
\boldsymbol{q}^{k+1} &= \boldsymbol{q}^k+\frac{\Delta t}{2}\hat{\boldsymbol{p}}^k\\
\begin{pmatrix}\boldsymbol{p}^{k+1}\\ \boldsymbol{s}^{k+1}\end{pmatrix} &= -\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}_{m+n}-\boldsymbol{F}_{\Delta t/2})\nabla_{\boldsymbol{q}}U(\boldsymbol{q}^{k+1})+\boldsymbol{F}_{\Delta t/2}\begin{pmatrix}\hat{\boldsymbol{p}}^k\\ \hat{\boldsymbol{s}}^k\end{pmatrix}+\boldsymbol{S}_{\Delta t/2}\hat{\mathscr{R}}^k
\end{aligned} \tag{B.3}$$

gle-ASA

$$\begin{aligned}
\boldsymbol{q}^{k+1/2} &= \boldsymbol{q}^k+\frac{\Delta t}{2}\boldsymbol{p}^k\\
\begin{pmatrix}\boldsymbol{p}^{k+1}\\ \boldsymbol{s}^{k+1}\end{pmatrix} &= -\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}_{m+n}-\boldsymbol{F}_{\Delta t})\nabla_{\boldsymbol{q}}U(\boldsymbol{q}^{k+1/2})+\boldsymbol{F}_{\Delta t}\begin{pmatrix}\boldsymbol{p}^k\\ \boldsymbol{s}^k\end{pmatrix}+\boldsymbol{S}_{\Delta t\Delta t}\mathscr{R}^k\\
\boldsymbol{q}^{k+1} &= \boldsymbol{q}^{k+1/2}+\frac{\Delta t}{2}\boldsymbol{p}^{k+1}
\end{aligned} \tag{B.4}$$

gle-BB3

$$\begin{aligned}
\boldsymbol{p}_i^{n+1/2} &= \boldsymbol{p}_i^n-\frac{\Delta t}{2}\nabla_{\boldsymbol{q}_i}U(\boldsymbol{q}_i^n)+\frac{\Delta t}{2}\sum_{k=1}^m \boldsymbol{s}_k^n\\
\boldsymbol{q}_i^{n+1} &= \boldsymbol{q}_i^n+\Delta t m^{-1}\boldsymbol{p}_i^{n+1/2}\\
\boldsymbol{s}_k^{n+1} &= \theta_k\boldsymbol{s}_k^n-(1-\theta_k)c_k\boldsymbol{p}_i^{n+1/2}+\alpha_k\sqrt{2\beta^{-1}c_k}\mathscr{R}_k^n\\
\boldsymbol{p}_i^{n+1} &= \boldsymbol{p}_i^{n+1/2}-\frac{\Delta t}{2}\nabla_{\boldsymbol{q}_i}U(\boldsymbol{q}_i^{n+1})+\frac{\Delta t}{2}\sum_{k=1}^m \boldsymbol{s}_k^{n+1}
\end{aligned} \tag{B.5}$$

where $\theta_k=e^{-\Delta t/\tau_k}$, $\alpha_k=\sqrt{\frac{(1-\theta_k)^2}{\Delta t}}$ and $\mathscr{R}_k^n,\ \ 1\le k\le m,\ n\in\mathbb{N}$ are i.i.d. normal distributed random variables

gle-BB3b

$$\begin{aligned}
\boldsymbol{p}_i^{n+1/2} &= \boldsymbol{p}_i^n-\frac{\Delta t}{2}\nabla_{\boldsymbol{q}_i}U(\boldsymbol{q}^n)\\
\boldsymbol{q}_i^{n+1/2} &= \boldsymbol{q}_i^n+m^{-1}\frac{\Delta t}{2}\boldsymbol{p}_i^n\\
\tilde{\boldsymbol{p}}^{n+1/2} &= \boldsymbol{p}_i^{n+1/2}-\frac{\Delta t}{2}\sum_{i=1}^m \boldsymbol{s}_i^n\\
\boldsymbol{s}_k^{n+1} &= \theta_k\boldsymbol{s}_k^n-(1-\theta_k)c_k m^{-1}\tilde{\boldsymbol{p}}^{n+1/2}+\alpha_k\sqrt{2\beta^{-1}c_k}\mathscr{R}_k^n\\
\hat{\boldsymbol{p}}^{n+1/2} &= \tilde{\boldsymbol{p}}^{n+1/2}-\frac{\Delta t}{2}\sum_{i=1}^m \boldsymbol{s}_i^{n+1}\\
\boldsymbol{q}_i^{n+1} &= \boldsymbol{q}_i^{n+1/2}+\frac{\Delta t}{2}\hat{p}^{n+1/2}\\
\boldsymbol{p}_i^{n+1} &= \hat{p}^{n+1/2}-m^{-1}\frac{\Delta t}{2}\nabla_{\boldsymbol{q}_i}U(\boldsymbol{q}^{n+1/2})
\end{aligned} \tag{B.6}$$

where θ_k, α_k and $\mathscr{R}_k^n,\ \ 1\le k\le m,\ n\in\mathbb{N}$ are defined as in (B.5).

gle-SKL

$$
\begin{aligned}
\boldsymbol{s}_{i,1}^{n+1/2} &= \boldsymbol{s}_{i,1}^{n}\xi(\Delta t/2) + \zeta(\Delta t/2)\mathscr{R}_1^n, & \boldsymbol{s}_{i,2}^{n+1/2} &= \boldsymbol{s}_{i,2}^{n}\xi(\Delta t/2) + \zeta(\Delta t/2)\mathscr{R}_2^n,\\
\boldsymbol{p}_i^{n+1/2} &= \boldsymbol{p}_i^n - \frac{\Delta t}{2}\left(\nabla_{\boldsymbol{q}_i}U(\boldsymbol{q}^n) + \sqrt{c}\boldsymbol{s}_{i,1}^{n+1/2}\right), & \tilde{\boldsymbol{s}}_2^{n+1/2} &= \boldsymbol{s}_{i,2}^{n+1/2} + \frac{\Delta t}{2}\alpha\boldsymbol{s}_{i,1}^{n+1/2},\\
\boldsymbol{p}_i^{n+1} &= \boldsymbol{p}_i^n + \Delta t\boldsymbol{p}_i^n, & \tilde{\boldsymbol{s}}_{i,1}^{n+1/2} &= \boldsymbol{s}_{i,1}^{n+1/2} + \sqrt{c}\boldsymbol{p}_i^{n+1/2} - \alpha\tilde{\boldsymbol{s}}_{i,2}^{n+1/2},\\
\boldsymbol{p}_i^{n+1} &= \boldsymbol{p}_i^{n+1/2} - \frac{\Delta t}{2}\left(\nabla_{\boldsymbol{q}_i}U(\boldsymbol{q}^{n+1}) + \sqrt{c}\tilde{\boldsymbol{s}}_{i,1}^{n+1/2}\right), & \hat{\boldsymbol{s}}_{i,2}^{n+1/2} &= \boldsymbol{s}_{i,2}^{n+1/2} + \frac{\Delta t}{2}\alpha\tilde{\boldsymbol{s}}_{i,1}^{n+1/2},\\
\boldsymbol{s}_{i,1}^{n+1} &= \tilde{\boldsymbol{s}}_{i,1}^{n+1/2}\xi(\Delta t/2) + \zeta(\Delta t/2)\tilde{\mathscr{R}}_1^n, & \boldsymbol{s}_{i,2}^{n+1} &= \hat{\boldsymbol{s}}_{i,2}^{n+1/2}\xi(\Delta t/2) + \zeta(\Delta t/2)\tilde{\mathscr{R}}_2^n,
\end{aligned}
\tag{B.7}
$$

where $\xi(\Delta t) = e^{-\frac{\Delta t}{\tau}}$ and $\zeta(\Delta t) = \sqrt{\beta^{-1}(1 - e^{-\frac{2\Delta t}{\tau}})}$.

Appendix C

Large expressions

Definition of g_Γ

The following function is used in the derivation of the super-convergence results of gle-BAOAB in Section 4.6.2.

$$
\begin{aligned}
&g_\Gamma(q,p,s) = \\
&\frac{-2\left(s\Gamma_{2,1}-p\Gamma_{2,2}\right)\left(\Gamma_{2,2}^2\left(\left(\beta p^2+6\right)U^{(3)}(q)-72v'(q)\right)-2\beta sp\Gamma_{2,2}\Gamma_{2,1}U^{(3)}(q)+\Gamma_{2,1}^2\left(\beta s^2-3\right)U^{(3)}(q)\right)}{g(\Gamma)} \\
&+\frac{2\Gamma_{1,1}^2\left(p\Gamma_{2,2}\left(\left(\beta p^2+6\right)U^{(3)}(q)-72v'(q)\right)+9s\Gamma_{2,1}\left(8v'(q)-U^{(3)}(q)\right)\right)}{g(\Gamma)} \\
&+\frac{\Gamma_{1,1}\left(\Gamma_{2,1}\left(3s\Gamma_{2,2}\left(120v'(q)-\left(\beta p^2+14\right)U^{(3)}(q)\right)-2p\Gamma_{1,2}\left(\beta p^2-3\right)U^{(3)}(q)\right)\right)}{g(\Gamma)} \\
&+\frac{\Gamma_{1,1}5p\Gamma_{2,2}^2\left(\left(\beta p^2+6\right)U^{(3)}(q)-72v'(q)\right)}{g(\Gamma)} \\
&+\frac{\Gamma_{1,2}\Gamma_{2,1}\left(3s\Gamma_{2,1}\left(\left(\beta p^2+2\right)U^{(3)}(q)-24v'(q)\right)+p\Gamma_{2,2}\left(\left(6-5\beta p^2\right)U^{(3)}(q)+72v'(q)\right)\right)}{g(\Gamma)}
\end{aligned}
$$

with

$$
g(\Gamma) = 72\left(\Gamma_{1,2}\Gamma_{2,1}-\Gamma_{1,1}\Gamma_{2,2}\right)\left(-2\Gamma_{1,1}^2-5\Gamma_{2,2}\Gamma_{1,1}-2\Gamma_{2,2}^2+\Gamma_{1,2}\Gamma_{2,1}\right).
$$

Parametrisation of pre-optimised memory kernel

The matrix representation of the memory kernel used in the numerical experiment described in Section 4.8.4 is of the form

$$
\boldsymbol{\Gamma}_{\text{kv\_8\_8}} := \boldsymbol{I}_n \otimes \begin{pmatrix} \Gamma_{1,1} & \Gamma_{1,2} \\ \Gamma_{2,1} & \Gamma_{2,2} \end{pmatrix},
$$

with $n = 9$, and

$$\Gamma_{1,1} = \big(\ 1.336001\text{E+}1\ \big),$$

$$\Gamma_{1,2} = \big(\ 8.327012\text{E-}6\ 1.850437\text{E-}4\ 2.551111\text{E-}3\ -1.63314\text{E-}2 -1.334317\text{E-}1\ 1.679873\text{E+}0\ 2.22050\text{E+}1\ 6.274743\text{E+}0\ \big),$$

$$\Gamma_{2,1} = \big(\ 1.20137\text{E-}5\ 1.83376\text{E-}4\ 2.505216\text{E-}3\ -1.625490\text{E-}2\ -1.334317\ 1.68179\text{E+}0\ 2.22086\text{E+}1\ 6.09239\text{E+}0\ \big)^T,$$

$$\Gamma_{2,2} = \begin{pmatrix} 3.25571\text{E-}6 & 7.47982\text{E-}6 & 9.622039\text{E-}6 & 4.244713\text{E-}5 & 1.695553\text{E-}5 & 3.285529\text{E-}5 & 6.747855\text{E-}6 & -1.438594\text{E-}4 \\ -7.479820\text{E-}6 & 1.019983\text{E-}4 & 1.424663\text{E-}5 & -1.396013\text{E-}5 & 1.513710\text{E-}5 & -1.200000\text{E-}5 & 2.54657\text{E-}5 & -6.346355\text{E-}5 \\ -9.622039\text{E-}6 & -1.424663\text{E-}5 & 2.206513\text{E-}3 & 2.2645018\text{E-}5 & 1.384732\text{E-}5 & 4.388201\text{E-}5 & -4.079418\text{E-}6 & 8.9663528\text{E-}3 \\ -4.244713\text{E-}5 & 1.396013\text{E-}5 & -2.26450\text{E-}5 & 2.218067\text{E-}2 & 1.249881\text{E-}5 & 2.492691\text{E-}5 & 1.116974336243\text{E-}5 & 3.14859310\text{E-}3 \\ -1.6955539\text{E-}5 & -1.513710\text{E-}5 & -1.3847329\text{E-}5 & -1.249881\text{E-}5 & 1.772222\text{E-}1 & 3.888513\text{E-}5 & -4.4198267\text{E-}6 & 1.010943\text{E-}1 \\ -3.285529\text{E-}5 & 1.200000\text{E-}5 & -4.388201\text{E-}5 & -2.492691\text{E-}5 & -3.888513\text{E-}5 & 2.79177\text{E+}0 & 8.375329164851\text{E-}6 & 2.17612\text{E-}1 \\ -6.74785\text{E-}6 & -2.54657\text{E-}5 & 4.07941\text{E-}6 & -1.116974\text{E-}5 & 4.419826\text{E-}6 & -8.37532\text{E-}6 & 4.043272\text{E+}1 & 3.460867415178\text{E-}2 \\ 1.438594\text{E-}4 & 6.346355\text{E-}5 & -8.966352\text{E-}3 & -3.14859\text{E-}3 & -1.010943\text{E-}1 & -2.176129826302\text{E-}1 & -3.46086\text{E-}2 & 1.088614\text{E+}3 \end{pmatrix}.$$